INTERACTION DESIGN FOUNDATION

# ENCYCLOPEDIA

OF HUMAN-COMPUTER INTERACTION

INTERACTION DESIGN FOUNDATION
Est. 2002

INTERACTION DESIGN FOUNDATION

# ENCYCLOPEDIA

OF HUMAN-COMPUTER INTERACTION

3RD EDITION, DRAFT

INTERACTION DESIGN FOUNDATION
Est. 2002

Title: The Encyclopedia of Human-Computer Interaction, 2nd Ed.
Authors: Mads Soegaard and Rikke Friis Dam
Publisher: The Interaction Design Foundation
Edition number: 1
ISBN paperback edition:
ISBN online edition: 978-87-92964-00-7
ISBN eBook edition:
ISBN hardback edition:
Copy editor: Ailsa Campbell
Reference validation: Armin Walinsky
Graphics, interior design, and cover: Roland Szattler and Mads Soegaard
Copyright clearance: Michael Thorman
Proofreader: Ailsa Campbell
Supervisory Editors: Mads Soegaard and Rikke Friis Dam
Editorial assistant: Soeren Dam
Typography: This text is set in Georgia
Copyright: See section on Copyright
Front page cover:
Acknowledgement:
Peer review: This book has undergone double-blinded peer-review based on the reviewing guidelines

Sponsored by

Production and dissemination of this work has been funded in part by SAP.

# Open Access

We believe in Open Access and the **democratization of knowledge**. Unfortunately, world class educational materials are normally hidden behind payment systems or in expensive textbooks. If you want this to change, you should help us out! Kind thoughts are **not** enough - you need to act!

# Table of Contents

# Chapter 1

# Interaction Design

## *by Jonas Lowgren.*

The aim of the following chapter is to provide an introductory overview of the concept and the field of interaction design, loosely grounded in historical developments. This encyclopedia covers the full gamut of human-computer interaction (HCI), and it should be noted that interaction design covers only a part of the HCI field. My intention here is to provide a frame of reference that can be used in reading other, more substantial chapters to start filling the notion of interaction design with solid topical content. This chapter itself is brief and superficial, paints with a broad brush; yet it is my hope that it conveys some of the key characteristics and considerations of interaction design, thus informing the reading of the topical chapters.

In his 2007 book *Designing Interactions*, industrial designer and IDEO founder Bill Moggridge reminisces (p. 14):

*"I felt that there was an opportunity to create a new design discipline, dedicated to creating imaginative and attractive solutions in a virtual world, where one could design behaviors, animations, and sounds as well as shapes. This would be the equivalent of industrial design but in software rather than three-dimensional objects. Like industrial design, the discipline would start from the needs and desires of the people who use a product or service, and strive to create designs that would give aesthetic pleasure as well as lasting satisfaction and enjoyment.*

*I gave my first conference presentation on the subject in 1984, and at that time I described it as "Soft-face", thinking of a combination between software and user-interface design [...] we went on thinking of possible names until I eventually settled on "interaction design" with the help of Bill Verplank."*

*-- Moggridge, 2007*

The interaction design label remained relatively marginal until the mid-1990s; the design community largely considered the behaviors of the virtual world to be a specialty within industrial design. During this period, academia as well as ICT industries were mainly occupied with usability and human factors engineering, focusing on ways to operationalize psychology and ergonomics into methods for creating efficient and error-free interactions to support work tasks.

## 1.1 FIVE MAJOR CHARACTERISTICS OF INTERACTION DESIGN

With the increasing penetration of the Internet, the advent of home and leisure computing, and eventually the emergence of digital interactive consumer products, the two cultures of design and engineering gravitated towards a common interest in discretionary use and user experience. Towards the turn of the century, the notion of interaction design started to gain in popularity as a way to acknowledge a more designerly approach to the topic – going beyond pure utility and efficiency to consider also aesthetic qualities of use, for example.

Since then, a plethora of professional practices, academic study programs, literatures, networks and venues have formed under the umbrella of interaction design. It goes without saying that there are many different understandings of exactly what interaction design is. I don't see any real point in surveying all these definitions but instead I would like to offer a very simple formulation of interaction design, devised to capture the heritage of the term as outlined above and at the same time draw some demarcation lines to indicate potential edges of the field. It goes like this:

*"Interaction design is about shaping digital things for people's use"*

This is indeed a simple formulation. However, as we shall see in the following where I discuss one of its elements at a time, it is not entirely without power of discrimination.

The notion of **shaping** is used consciously to suggest a designerly activity (as opposed to, e.g., "building" which suggests engineering, or "making" or "creating" that could refer to more or less anything). More specifically, I find it to be a distinctive trait of interaction design that the gestation process is a Design process, in the capital-D sense of the word. This in turn implies five major characteristics.

### 1.1.1 Design involves changing situations by shaping and deploying artifacts

In other words, design is about transformation and the means available for the designer to initiate change in a particular situation is ultimately the designed artifact.

For interaction design, this connects to the notion of what the interaction designer designs. I am suggesting the delimitation that interaction designers design digital things – more on this below. What that means for now, however, is that changing a situation by devising and implementing, say, a new political initiative could certainly be viewed as a design act but not an act of interaction design.

### 1.1.2 Design is about exploring possible futures

This seems almost too obvious to point out, but from an academic point of view it might be worth mentioning since it entails a fundamental difference in orientation; analytical and critical studies focus on that which exists, whereas design concerns itself with that which could be. This has epistemological consequences for, e.g., how research is conducted. Framing design as exploration also means that it often makes sense to spend time in early phases on divergent work, essentially looking around in a design space of possibilities before committing to a particular direction. Exploring possible futures in interaction design often involves inviting the future users in various forms of participation.

Claiming that design entails exploring possible futures also means that activities like user studies and summative evaluations in themselves do not constitute interaction design. However, they are often used within interaction design processes, and arguably it makes sense to consider the larger process including fieldwork, innovation and evaluation as a design process in its entirety – as the larger process is actually about exploring possible futures.

### 1.1.3 Design entails framing the "problem" in parallel with creating possible "solutions"

From the notions of changing situations and exploring possible futures follow the conclusion that when we have designed something, the situation in which it is used is no longer the same. This in turn means that analyzing the existing in order to define a "problem" – that subsequent design should solve – is essentially of limited merit. Exploring possible futures implies not only different "design solutions" but also different "problems." For contemporary interaction design practice, this has implications such as reconsidering notions of exhaustive specifications before build in favor of perpetual-beta approaches and the like.

A consequence of this characteristic is that traditional systems development and engineering processes, where the aim is to finish descriptive analysis for a requirement specification before creative design begins, are not considered designerly processes. This is quite intentional.

### 1.1.4 Design involves thinking through sketching and other tangible representations

When sketching snapshots or aspects of possible futures (such as a not-yet-existing product), the designer is not merely copying images from her inner eye. The drawings are micro-experiments that respond with insights into strengths, weaknesses and possible changes in a tight loop of thinking that involves the hand, the senses and the mind. The same notion applies for other sketching media used in design practice. For interaction design, there are particular implications to be observed from the temporal nature of our design material. One of them is that when designing innovative interaction techniques, it may be necessary to sketch in software and hardware rather than staying with lo-fi sketching media.

In general, the notion of sketching is more about the mindset of the designer than about the medium used. If a particular external representation serves to engage the designer in a conversation about the details and implications of a not-yet-finalized idea, and if it is quick, tentative and truly disposable, then it is a sketch. It could be anything from a napkin drawing to a piece of programming code, perhaps even written in the language that is normally used to build products for delivery – what matters is the purpose and intention.

### 1.1.5 Design addresses instrumental, technical, aesthetical and ethical aspects throughout

Each of the possible futures being explored in a design process introduces considerations and tradeoffs in all these dimensions, and there is no obvious way in which they can be sequenced. This holds equally for interaction design: Technical decisions influence the aesthetic qualities of the resulting interaction, instrumental choices on features to offer have ethical repercussions, and so on.

Historically, there has been a tendency in human-computer interaction, usability engineering and human factors to focus on instrumental and technical aspects. Interaction design as a designerly activity would insist that the aesthetical and ethical qualities can never be ignored or factored out. Whether something looks and feels good to use, and whether it makes you comfortable in terms of social accountability and moral standards, has a real impact not only on the overall user experience but also on measurable, instrumental outcomes. For an interaction designer, users are whole people with complex sensibilities and design processes need to be conducted accordingly.

## 1.2 DIGITAL MATERIALS AND INTERACTION DESIGN

**Digital things** are what interaction design shapes. This is essentially to say that interaction designers work in *digital materials* – software, electronics, communication networks, and the like. And, as pointed out above, the digital materials pose specific requirements on, e.g., sketching practices. When designing an innovative interaction technique, where there is not much previous experience to rely on, it is sometimes necessary to experiment with constructions in software and/or hardware. Those constructions should be made with a sketching mindset, however, which among other things means that it is quickly made, focuses on behaviors and effects, is disposable and ideally also that it is one among many variations on the same theme (see above).

Historically, the digital things made by interaction designers were largely *tools* – contraptions intended to be used instrumentally, for solving problems and carrying out tasks, and mostly to be used individually. Much of our ingrained best-practice knowledge in the field emanates from this time, expressed in concepts such as user goals, task flows, usability and utility. However, it turns out that digital technology in society today is mostly used for communication, i.e., as a *medium*. And as a medium, it has characteristics that set it apart from previously existing personal and mass communication media. For example, it lowers the thresholds of media production to include virtually anyone, it provides many-to-many communication with persistent records of all exchanges that transpire, and it offers access to ongoing modifications of its infrastructures. These characteristics of what we might call *collaborative media* are only beginning to be understood in interaction design, and one might expect that this will be one of the most significant areas for future conceptual developments in our field.

By limiting the scope of interaction design to digital things (including media), we also *exclude* large parts of service design, organizational design, sociopolitical intervention, and so on. A historical analogy may be the typical experience of an enterprise systems consultant in the 1980s whose client asked for a new system to manage payroll. Analyzing the current situation might have turned up the insight that the old system as such had no major shortcomings, but that the workflow of the personnel department was severely convoluted and crippled. Would the consultant propose a new system anyway, or more rightly point out the need for an organizational development consultant? Or perhaps even try her own hand on organizational intervention?

Similar situations are legion in contemporary interaction design, as the use of digital technology is often deeply intertwined with other aspects of everyday life in the design situations approached by the interaction designer. What I propose – that interaction design creates digital things – should be understood as a recognition of the complexities and professional demands involved in related disciplines such as service design, urban development and political change. Essentially, the position adopted here is that when an interaction design process moves into the territory of non-digital intervention, the ideal scenario would see the establishment of a *multidisciplinary* design team. In practical work, however, this is not always a feasible option. The short-term benefits of being able to deliver must then be weighed against the potential long-term risks of doing a less-than-professional job in a related field.

## 1.3 PEOPLE'S USE AND INTERACTION DESIGN

**People's use** is what interaction design shapes digital things for. As indicated above, the historical notion of people's use was tightly connected to workplace settings and instrumental motivations: Use the program to get the job done as quickly, efficiently and correctly as possible. With the growth of digital technology outside the workplace in the form of consumer products came other notions of use, such as using for entertainment and for pleasure. Internet penetration has made way for use as communication, which is arguably today the most prominent kind of use of digital technology.

This broadened understanding of use has had a major impact on interaction design, most notably in the rise of the notion of *user experience* to capture all manners of non-instrumental, aesthetical, emotional qualities in the human use of a digital thing. However, following on from the heritage of digital things as individual tools, user experience in the literature is mostly an individual construct. Qualities that are essential social or communal in their nature, such as ethical implications and aspects of communication, are as yet somewhat underdeveloped in interaction design. Again, with the development of digital things towards collaborative media, one might expect more interest in this area in the near future.

**To conclude,** interaction design can be understood as shaping digital things for people's use. The practice of interaction design is knowledge-intensive and multidisciplinary at heart. The chapters of this encyclopedia provide much of the relevant knowledge that forms the basis for interaction design practice as well as its scholarship.

## 1.4 WHERE TO LEARN MORE

To me, the most approachable book-length introduction to interaction design is *Designing for interaction: Creating innovative applications and devices* by Dan Saffer (New Riders, 2nd ed., 2009).

Following on from that, *Sketching the user experience: Getting the design right and the right design* by Bill Buxton (Morgan Kaufmann, 2007) offers a very useful treatment of what a designerly approach to the digital materials means and what its implications are in the contemporary ICT industry.

Compared to other design fields, interaction design largely lacks a sense of a historical canon of products, concepts and designers. This is where *Designing interactions* by Bill Moggridge (MIT Press, 2007) comes in. It is an admirable first step towards establishing the much-needed discourse of the interaction design canon, and has a lot to offer for someone learning the field.

The book *Thoughtful interaction design: A design perspective on information technology* by myself and Erik Stolterman (MIT Press, 2004) introduces a number of concepts for thinking about interaction design processes, skills and practices.

A more extensive annotated bibliography of books pertinent to interaction design can be found at http://www.librarything.com/catalog/jonas.lowgren

The most significant professional network for interaction design is the Interaction Design Association (IXDA), which engages several thousands of interaction designers worldwide. The website at www.ixda.org offers several resources for professional learning and development, including a lively discussion forum. They also organize an annual international conference called Interaction.

Academic research in interaction design is somewhat scattered across venues. The premiere international conference on human-computer interaction is called CHI and is organized annually by ACM since the early 1980s. Its proceedings contain quite a lot of quality interaction-design research, as well as other work that is not as designerly in terms of approach and significance. The ACM also runs a smaller biannual conference called DIS (Designing Interactive Systems) that is more closely limited to interaction design. Moreover, there is a whole range of conferences in related fields where the interaction design student can find relevant material, such as Ubicomp and DPPI (Designing Pleasurable Products and Interfaces).

The academic field of HCI has a broad range of archival journals, where interaction-design research is occasionally published. Examples include *Human-Computer Interaction*, *ACM Transactions on Computer-Human Interaction* and *ACM Computers in Entertainment*. Finally, the magazine called *interactions* from ACM publishes many interaction-design related articles that aim to address professional as well as academic audiences.

The field of design research in general has less of an academic heritage than the field of HCI, and it comes as no surprise that its selection of academic literature is more limited. A notable exception is the *International Journal of Design*, which has quickly reached a respectable level of academic quality and which publishes interaction-design articles occasionally. Other journals that might be interesting for students of interaction design are *Design Issues* and *Digital Creativity*.

## 1.5 REFERENCES

Buxton, Bill (2007a): *Sketching User Experiences: Getting the Design Right and the Right Design*. Morgan Kaufmann

Buxton, Bill (2007b): *Sketching User Experiences: Getting the Design Right and the Right Design (Interactive Technologies)*. Morgan Kaufmann

Lowgren, Jonas and Stolterman, Erik A. (2004): *Thoughtful Interaction Design: A Design Perspective on Information Technology*. MIT Press

Moggridge, Bill (2007): *Designing Interactions*. The MIT Press

Saffer, Dan (2006): *Designing for Interaction: Creating Smart Applications and Clever Devices*. New Riders Press

Sharp, Helen, Rogers, Yvonne and Preece, Jennifer J. (2007): *Interaction Design: Beyond Human-Computer Interaction*. John Wiley and Sons

Winograd, Terry (1996): *Bringing Design to Software*. ACM Press

## About the author

### Jonas Lowgren

Has also published under the name of:
"Jonas Löwgren"

Personal Homepage:
http://webzone.k3.mah.se/k3jolo

Current place of employment:
Malmö University, Sweden

Jonas Löwgren is an interaction designer, researcher and teacher. Currently employed as professor of interaction design at Malmö University, Sweden. Main areas of expertise include cross-media products, interactive visualization and the design theory of digital materials

## YOUR NOTES AND THOUGHTS ON CHAPTER 1

Record your notes and thoughts on this chapter. If you want to share these thoughts with others online, go to the bottom of the page at:
http://www.interaction-design.org/encyclopedia/interaction_design.html

**NOTES:**

# Chapter 2

## Human Computer Interaction (HCI)

*by John M. Carroll.*

Human-computer interaction (HCI) is an area of research and practice that emerged in the early 1980s, initially as a specialty area in computer science embracing cognitive science and human factors engineering. HCI has expanded rapidly and steadily for three decades, attracting professionals from many other disciplines and incorporating diverse concepts and approaches. To a considerable extent, HCI now aggregates a collection of semi-autonomous fields of research and practice in human-centered informatics. However, the continuing synthesis of disparate conceptions and approaches to science and practice in HCI has produced a dramatic example of how different epistemologies and paradigms can be reconciled and integrated in a vibrant and productive intellectual project.

## 2.1 WHERE HCI CAME FROM

Until the late 1970s, the only humans who interacted with computers were information technology professionals and dedicated hobbyists. This changed disruptively with the emergence of personal computing in the later 1970s. Personal computing, including both personal software (productivity applications, such as text editors and spreadsheets, and interactive computer games) and personal computer platforms (operating systems, programming languages, and hardware), made everyone in the world a potential computer user, and vividly highlighted the deficiencies of computers with respect to *usability* for those who wanted to use computers as tools.

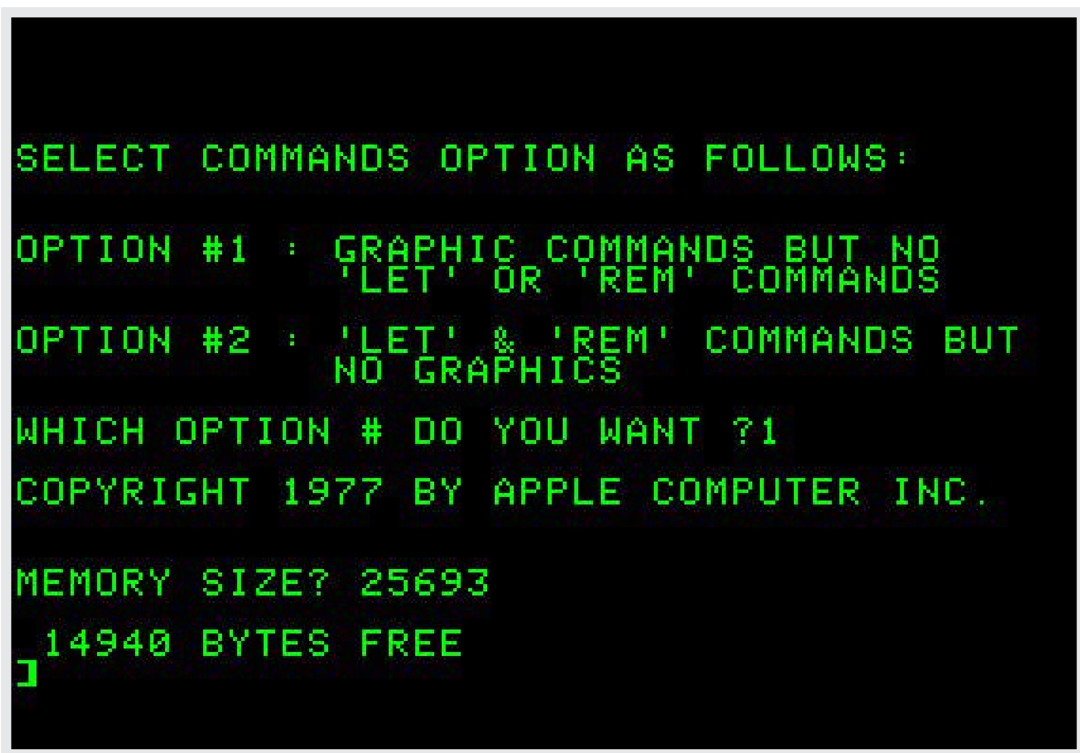





Figure 2.1 A-B: Personal computing rapidly pushed computer use into the general population, starting in the later 1970s. However, the non-professional computer user was often subjected to arcane commands and system dialogs.

The challenge of personal computing became manifest at an opportune time. The broad project of cognitive science, which incorporated cognitive psychology, artificial intelligence, linguistics, cognitive anthropology, and the philosophy of mind, had formed at the end of the 1970s. Part of the programme of cognitive science was to articulate systematic and scientifically informed applications to be known as "cognitive engineering". Thus, at just the point when personal computing presented the practical need for HCI, cognitive science presented people, concepts, skills, and a vision for addressing such needs through an ambitious synthesis of science and engineering. HCI was one of the first examples of cognitive engineering.

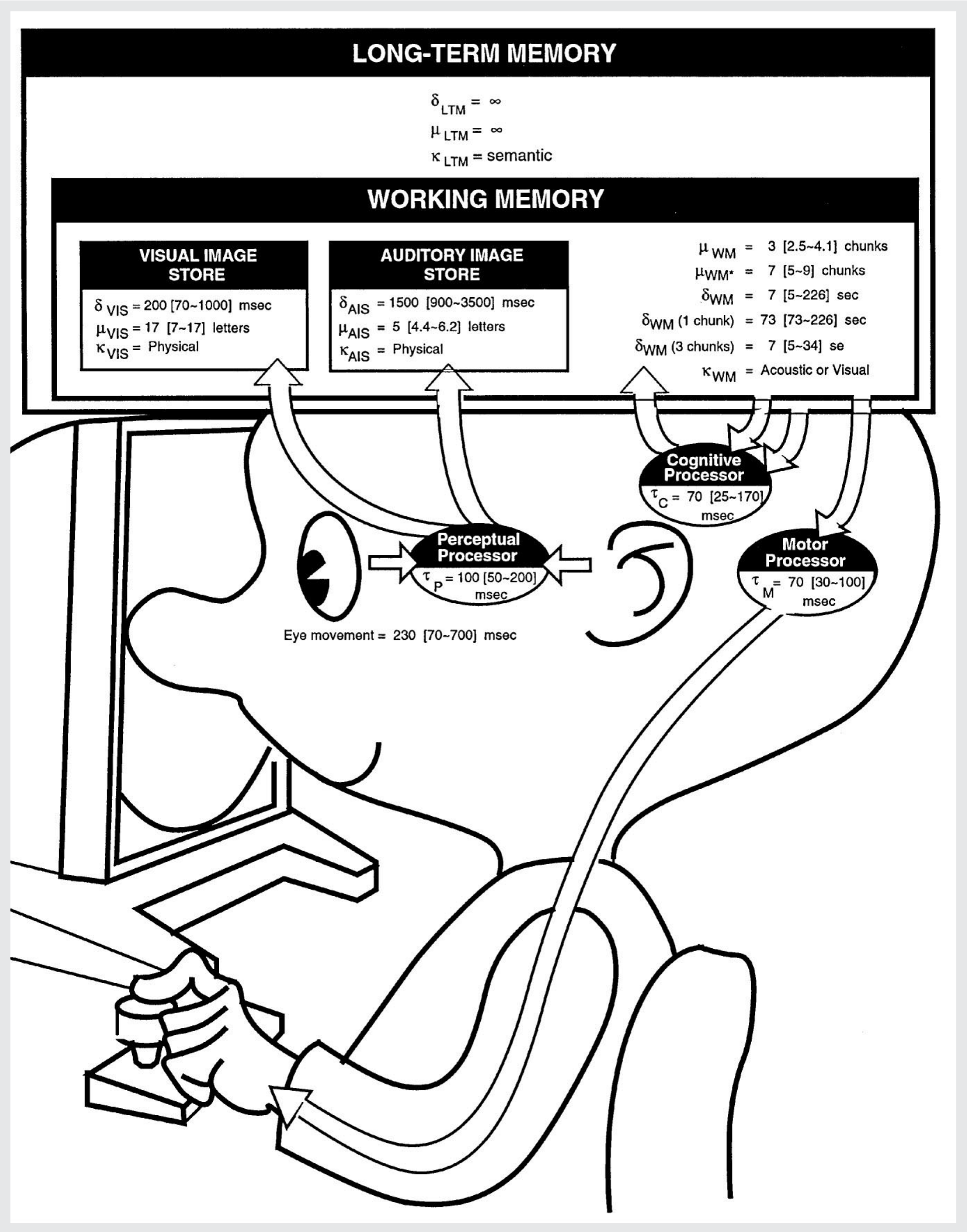



FIGURE 2.2: The Model Human Processor was an early cognitive engineering model intended to help developers apply principles from cognitive psychology.

This was facilitated by analogous developments in engineering and design areas adjacent to HCI, and in fact often overlapping HCI, notably human factors engineering and documentation development. Human factors had developed empirical and task-analytic techniques for evaluating human-system interactions in domains such as aviation and manufacturing, and was moving to address interactive system contexts in which human operators regularly exerted greater problem-solving discretion. Documentation development was moving beyond its traditional role of producing systematic technical descriptions toward a cognitive approach incorporating theories of writing, reading, and media, with empirical user testing. Documents and other information needed to be usable also.

MOVING THE CURSOR

The four cursor-movement keys have arrows on them (they are located on the right of the keyboard).

**PRESS THE ↓ CURSOR KEY SEVERAL TIMES AND WATCH THE CURSOR MOVE DOWN THE SCREEN.**

The ↑, ← , and → cursor keys work analogously. Try them and see.

If you move the cursor all the way to the bottom of the screen, or all the way to the right, the display "shifts" so that you can see more of your document. By moving the cursor al the way up and to the left, you can bring the document back to where it started.

DELETING TEXT

**USE THE CURSOR KEYS TO MOVE THE CURSOR UNDER THE FIRST r IN THE WORD regular.**

**PRESS THE DEL KEY**

The DEL key is located up and to the right of the keyboard keys. Is the Displaywriter prompting you?: **Delete what?**

► If you make a mistake at this point, use CODE + CANCL and start the deletion again.

**USING THE → KEY, MOVE THE CURSOR THROUGH THE MATERIAL TO BE DELETED, THE WORD regular.**

The word is highlighted: you can see exactly what is going to be deleted before it actually is deleted

► If the wrong characters are highlighted use CODE + CANCL and start the deletion again.

FIGURE 2.3: Minimalist information emphasized supporting goal-directed activity in a domain. Instead of topic hierarchies and structured practice, it emphasized succinct support for self-directed action and for recognizing and recovering from error.

Other historically fortuitous developments contributed to the establishment of HCI. Software engineering, mired in unmanageable software complexity in the 1970s (the “software crisis”), was starting to focus on nonfunctional requirements, including usability and maintainability, and on empirical software development processes that relied heavily on iterative prototyping and empirical testing. Computer graphics and information retrieval had emerged in the 1970s, and rapidly came to recognize that interactive systems were the key to progressing beyond early achievements. All these threads of development in computer science pointed to the same conclusion: The way forward for computing entailed understanding and better empowering users. These diverse forces of need and opportunity converged around 1980, focusing a huge burst of human energy, and creating a highly visible interdisciplinary project.

## 2.2 FROM CABAL TO COMMUNITY

The original and abiding technical focus of HCI was and is the concept of *usability*. This concept was originally articulated somewhat naively in the slogan “easy to learn, easy to use”. The blunt simplicity of this conceptualization gave HCI an edgy and prominent identity in computing. It served to hold the field together, and to help it influence computer science and technology development more broadly and effectively. However, inside HCI the concept of usability has been re-articulated and reconstructed almost continually, and has become increasingly rich and intriguingly problematic. Usability now often subsumes qualities like fun, well being, collective efficacy, aesthetic tension, enhanced creativity, flow, support for human development, and others. A more dynamic view of usability is one of a programmatic objective that should and will continue to develop as our ability to reach further toward it improves.

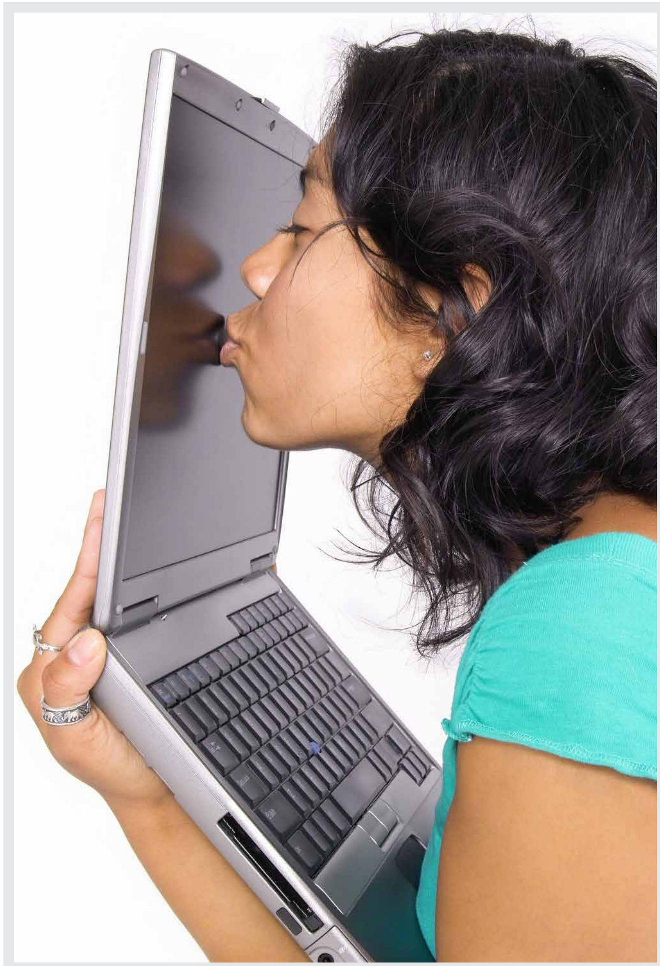


Figure 2.4: Usability is an emergent quality that reflects the grasp and the reach of HCI. Contemporary users want more from a system than merely "ease of use".



Although the original academic home for HCI was computer science, and its original focus was on personal productivity applications, mainly text editing and spreadsheets, the field has constantly diversified and outgrown all boundaries. It quickly expanded to encompass visualization, information systems, collaborative systems, the system development process, and many areas of design. HCI is taught now in many departments/faculties that address information technology, including psychology, design, communication studies, cognitive science, information science, science and technology studies, geographical sciences, management information systems, and industrial, manufacturing, and systems engineering. HCI research and practice draws upon and integrates all of these perspectives.

A result of this growth is that HCI is now less singularly focused with respect to core concepts and methods, problem areas and assumptions about infrastructures, applications, and types of users. Indeed, it no longer makes sense to regard HCI as a specialty of computer science; HCI has grown to be broader, larger and much more diverse than computer science itself. HCI expanded from its initial focus on individual and generic user behavior to include social and organizational computing, accessibility for the elderly, the cognitively and physically impaired, and for all people, and for the widest possible spectrum of human experiences and activities. It expanded from desktop office applications to include games, learning and education, commerce, health and medical applications, emergency planning and response, and systems to support collaboration and community. It expanded from early graphical user interfaces to include myriad interaction techniques and devices, multi-modal interactions, tool support for model-based user interface specification, and a host of emerging ubiquitous, handheld and context-aware interactions.

There is no unified concept of an HCI professional. In the 1980s, the cognitive science side of HCI was sometimes contrasted with the software tools and user interface side of HCI. The landscape of core HCI concepts and skills is far more differentiated and complex now. HCI academic programs train many different types of professionals: user experience designers, interaction designers, user interface designers, application designers, usability engineers, user interface developers, application developers, technical communicators/online information designers, and more. And indeed, many of the sub-communities of

HCI are themselves quite diverse. For example, ubiquitous computing (aka ubicomp) is subarea of HCI, but it is also a superordinate area integrating several distinguishable subareas, for example mobile computing, geo-spatial information systems, in-vehicle systems, community informatics, distributed systems, handhelds, wearable devices, ambient intelligence, sensor networks, and specialized views of usability evaluation, programming tools and techniques, and application infrastructures. The relationship between ubiquitous computing and HCI is paradigmatic: HCI is the name for a community of communities.

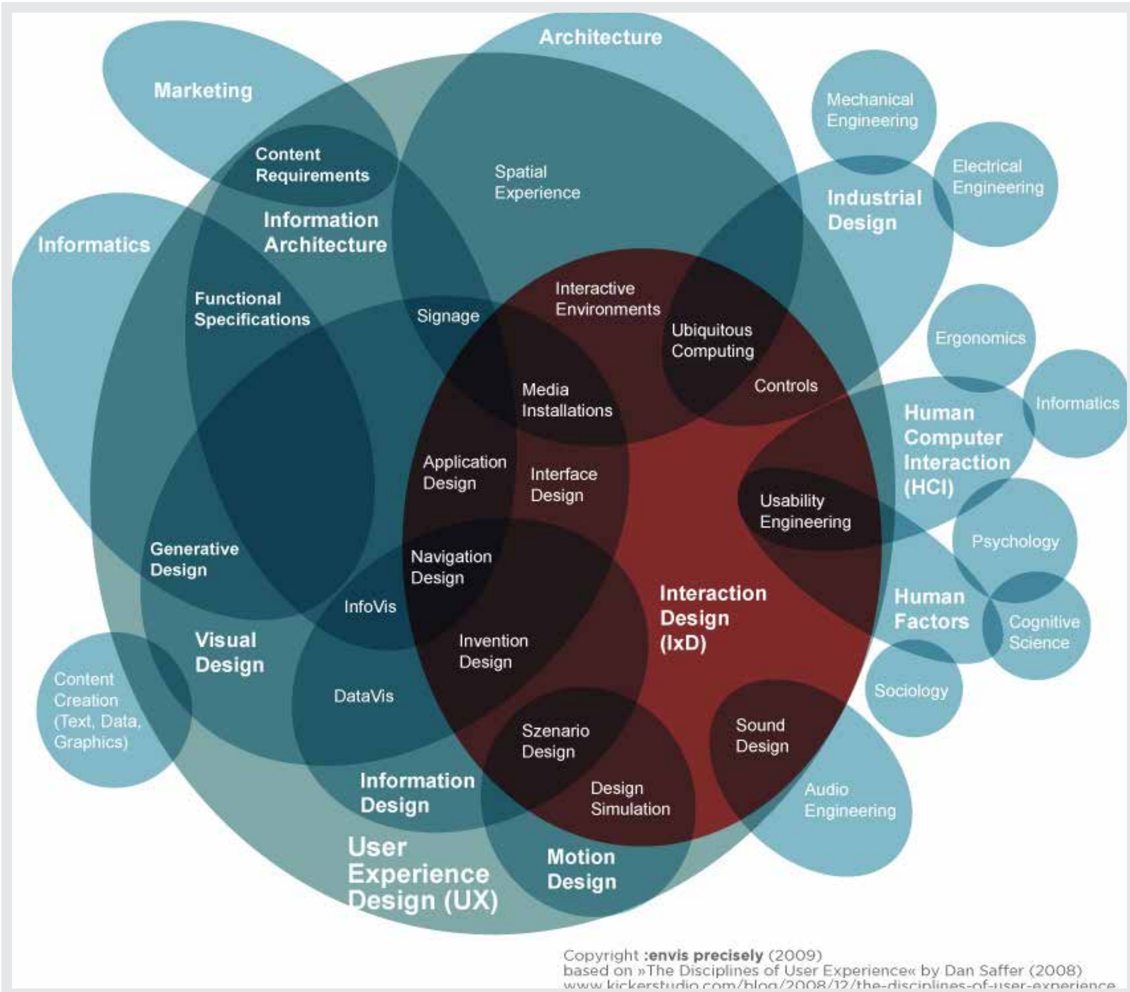



FIGURE 2.5 A-B: Two visualizations of the variety of disciplinary knowledge and skills involved in contemporary design of human-computer interactions.



Indeed, the principle that HCI is a community of communities is now a point of definition codified, for example, in the organization of major HCI conferences and journals. The integrating element across HCI communities continues to be a close linkage of critical analysis of usability, broadly understood, with development of novel technology and applications. This is the defining identity commitment of the HCI community. It has allowed HCI to successfully cultivate respect for the diversity of skills and concepts that underlie innovative technology development, and to regularly transcend disciplinary obstacles. In the early 1980s, HCI was a small and focused specialty area. It was a cabal trying to establish what was then a heretical view of computing. Today, HCI is a vast and multifaceted community, bound by the evolving concept of usability, and the integrating commitment to value human activity and experience as the primary driver in technology.

## 2.3 BEYOND THE DESKTOP

Given the contemporary shape of HCI, it is important to remember that its origins are personal productivity interactions bound to the desktop, such as word processing and spreadsheets. Indeed, one of biggest design ideas of the early 1980s was the so-called messy desk metaphor, popularized by the Apple Macintosh: Files and folders were displayed as icons that could be, and were scattered around the display surface. The messy desktop was a perfect incubator for the developing paradigm of graphical user interfaces. Perhaps it wasn't quite as easy to learn and easy to use as claimed, but people everywhere were soon double clicking, dragging windows and icons around their displays, and losing track of things on their desktop interfaces just as they did on their physical desktops. It was surely a stark contrast to the immediately prior teletype metaphor of Unix, in which all interactions were accomplished by typing commands.

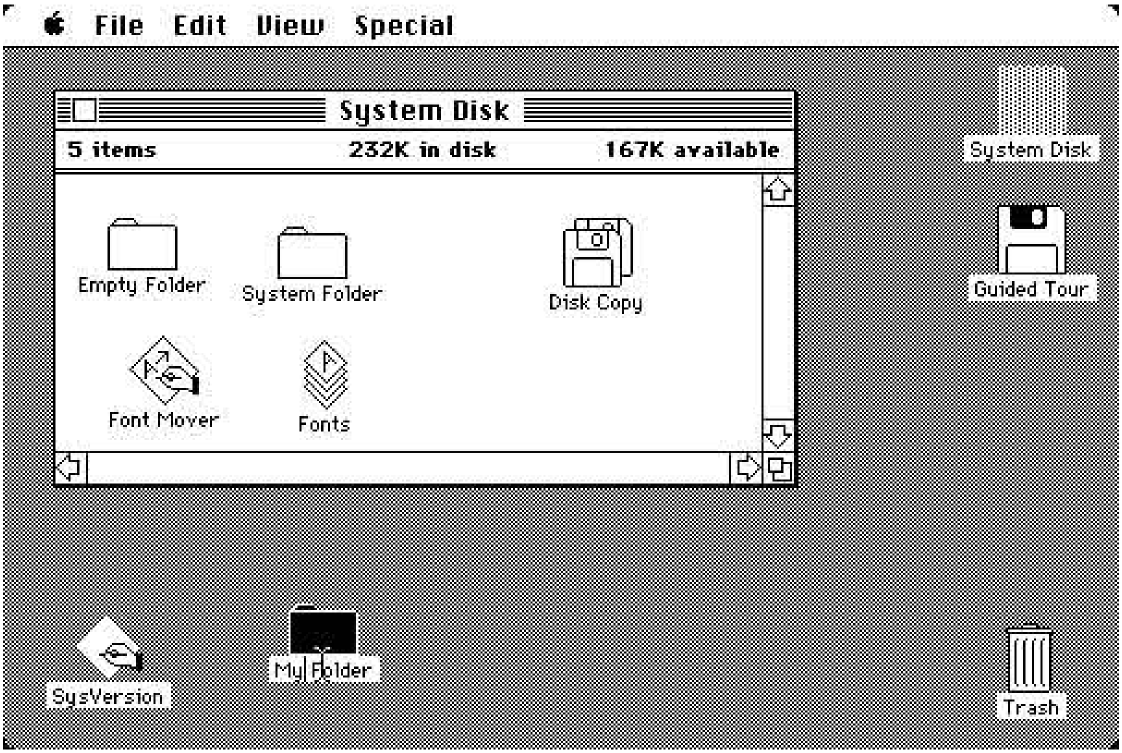



Figure 2.6: The early Macintosh desktop metaphor: Icons scattered on the desktop depict documents and functions, which can be selected and accessed (as System Disk in the example).



Even though it can definitely be argued that the desktop metaphor was superficial, or perhaps under-exploited as a design paradigm, it captured imaginations of designers and the public. These were new possibilities for many people in 1980, pundits speculated about how they might change office work. Indeed, the tsunami of desktop designs challenged, sometimes threatened the expertise and work practices of office workers. Today they are in the cultural background. Children learn these concepts and skills routinely.

As HCI developed, it moved beyond the desktop in three distinct senses. First, the desktop metaphor proved to be more limited than it first seemed. It's fine to directly represent a couple dozen digital objects as icons, but this approach quickly leads to clutter, and is not very useful for people with thousands of personal files and folders. Through the mid-1990s, HCI professionals and everyone else realized that search is a more fundamental paradigm than browsing for finding things in a user interface. Ironically though, when early World Wide Web pages emerged in the mid-1990s, they not only dropped the messy desktop metaphor, but for the most part dropped graphical interactions entirely. And still they were seen as a breakthrough in usability (of course, the direct contrast was to Unix-style tools like ftp and telnet). The design approach of displaying and directly interacting with data objects as icons has not disappeared, but it is no longer a hegemonic design concept.

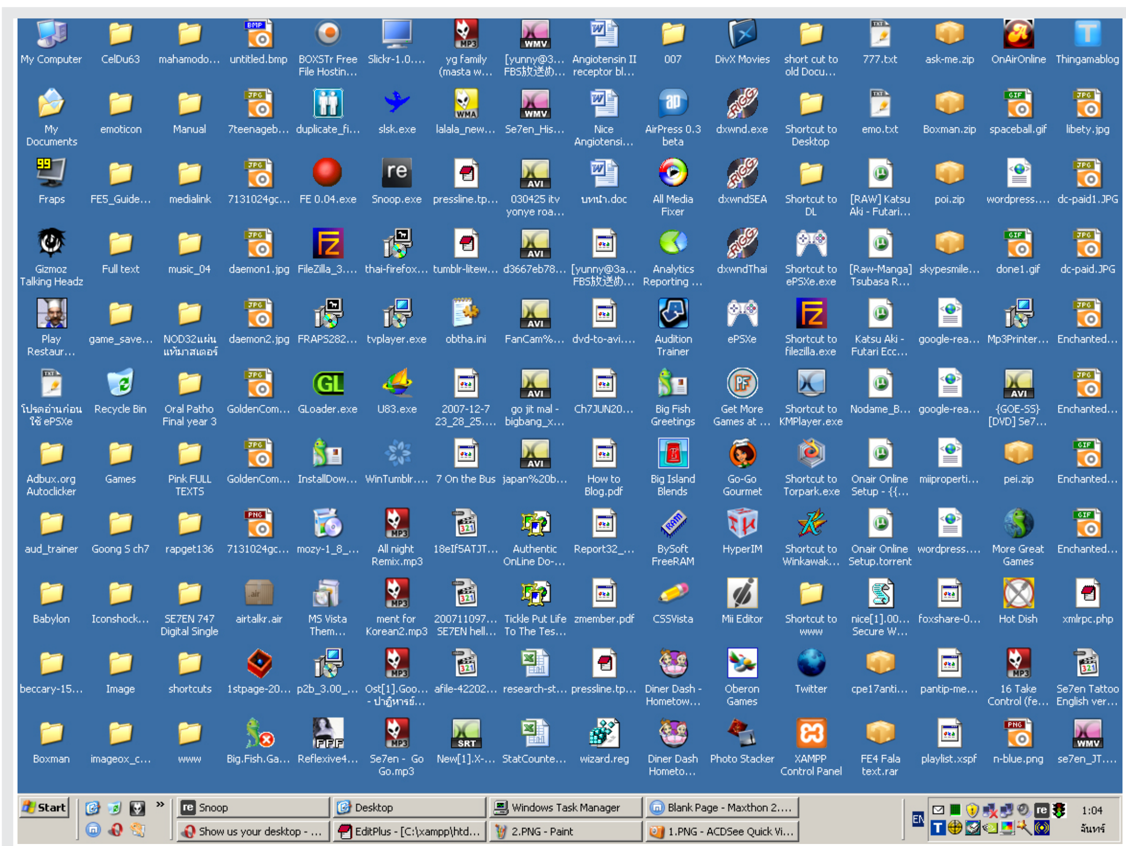


FIGURE 2.7: The early popularity of messy desktops for personal information spaces does not scale.

The second sense in which HCI moved beyond the desktop was through the growing influence of the Internet on computing and on society. Starting in the mid-1980s, email emerged as one of the most important HCI applications, but ironically, email made computers and networks into communication channels; people were not interacting *with* computers, they were interacting with other people *through* computers. Tools and applications to support collaborative activity now include instant messaging, wikis, blogs, online forums, social networking, social bookmarking and tagging services, media spaces and other collaborative workspaces, recommender and collaborative filtering systems, and a wide variety of online groups and communities. New paradigms and mechanisms for collective activity have emerged including online auctions, reputation systems, soft sensors, and crowd sourcing. This area of HCI, now often called social computing, is one of the most rapidly developing.

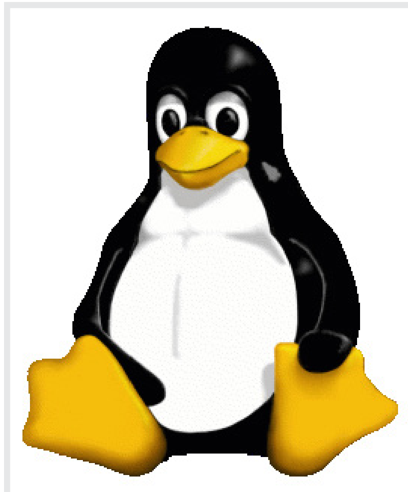







Figure 2.8 A-B-C: A huge and expanding variety of social network services are part of everyday computing experiences for many people. Online communities, such as Linux communities and GitHub, employ social computing to produce high-quality knowledge work.

The third way that HCI moved beyond the desktop was through the continual, and occasionally explosive diversification in the ecology of computing devices. Before desktop applications were consolidated, new kinds of device contexts emerged, notably laptops, which began to appear in the early 1980s, and handhelds, which began to appear in the mid-1980s. One frontier today is ubiquitous computing: The pervasive incorporation of computing into human habitats — cars, home appliances, furniture, clothing, and so forth. Desktop computing is still very important, though the desktop habitat has been transformed by the wide use of laptops. To a considerable extent, the desktop itself has moved off the desktop.

*Courtesy of Andrew Stern. Copyright: CC-Att-SA-3 (Creative Commons Attribution-ShareAlike 3.0).*

*Courtesy of United States Federal Government. Copyright: pd (Public Domain (information that is common property and contains no original authorship)).*


*Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.*

FIGURE 2.9 A-B-C: Computing moved off the desktop to be everywhere all the time. Computers are in phones, cars, meeting rooms, and coffee shops.

The focus of HCI has moved beyond the desktop, and its focus will continue to move. HCI is a technology area, and it is ineluctably driven to frontiers of technology and application possibility. The special value and contribution of HCI is that it will investigate, develop, and harness those new areas of possibility not merely as technologies or designs, but as means for enhancing human activity and experience.

## 2.4 THE TASK-ARTIFACT CYCLE

The movement of HCI off the desktop is a large-scale example of a pattern of technology development that is replicated throughout HCI at many levels of analysis. HCI addresses the dynamic co-evolution of the activities people engage in and experience, and the artifacts — such as interactive tools and environments — that mediate those activities. HCI is about understanding and critically evaluating the interactive technologies people use and experience. But it is also about how those interactions evolve as people appropriate technologies, as their expectations, concepts and skills develop, and as they articulate new needs, new interests, and new visions and agendas for interactive technology.

Reciprocally, HCI is about understanding contemporary human practices and aspirations, including how those activities are embodied, elaborated, but also perhaps limited by current infrastructures and tools. HCI is about understanding practices and activity specifically as requirements and design possibilities envisioning and bringing into being new technology, new tools and environments. It is about exploring design spaces, and realizing new systems and devices through the co-evolution of activity and artifacts, the task-artifact cycle.

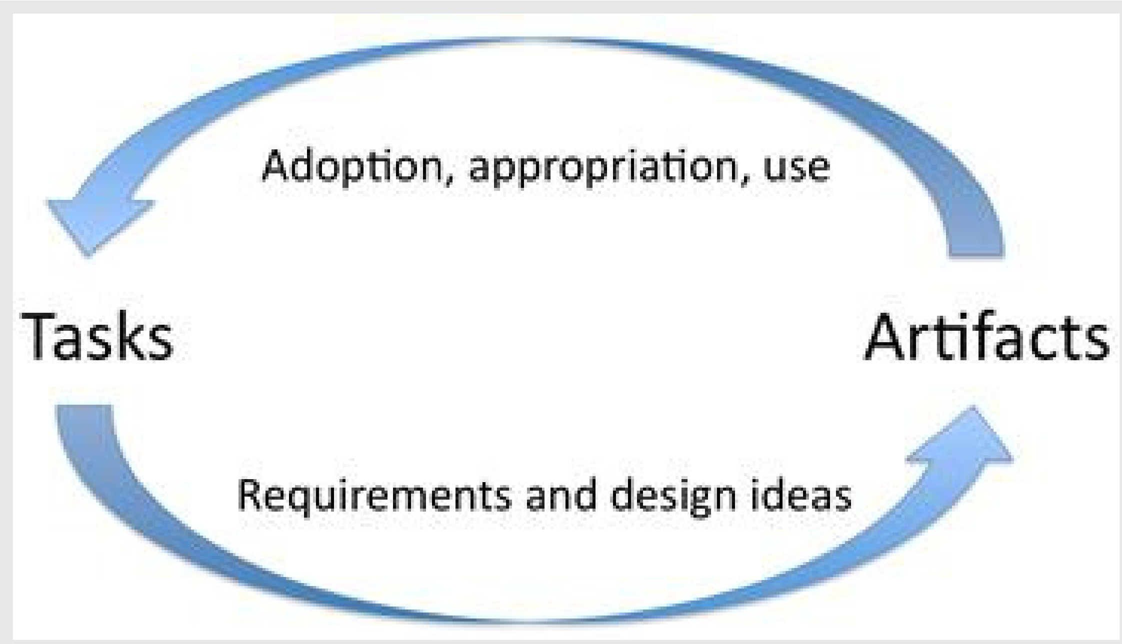



FIGURE 2.10: Human activities implicitly articulate needs, preferences and design visions. Artifacts are designed in response, but inevitably do more than merely respond. Through the course of their adoption and appropriation, new designs provide new possibilities for action and interaction. Ultimately, this activity articulates further human needs, preferences, and design visions.



Understanding HCI as inscribed in a co-evolution of activity and technological artifacts is useful. Most simply, it reminds us what HCI is like, that all of the infrastructure of HCI, including its concepts, methods, focal problems, and stirring successes will always be in flux. Moreover, because the co-evolution of activity and artifacts is shaped by a cascade of contingent initiatives across a diverse collection of actors, there is no reason to expect HCI to be convergent, or predictable. This is not to say progress in HCI is random or arbitrary, just that it is more like world history than it is like physics. One could see this quite optimistically: Individual and collective initiative shapes what HCI is, but not the laws of physics.

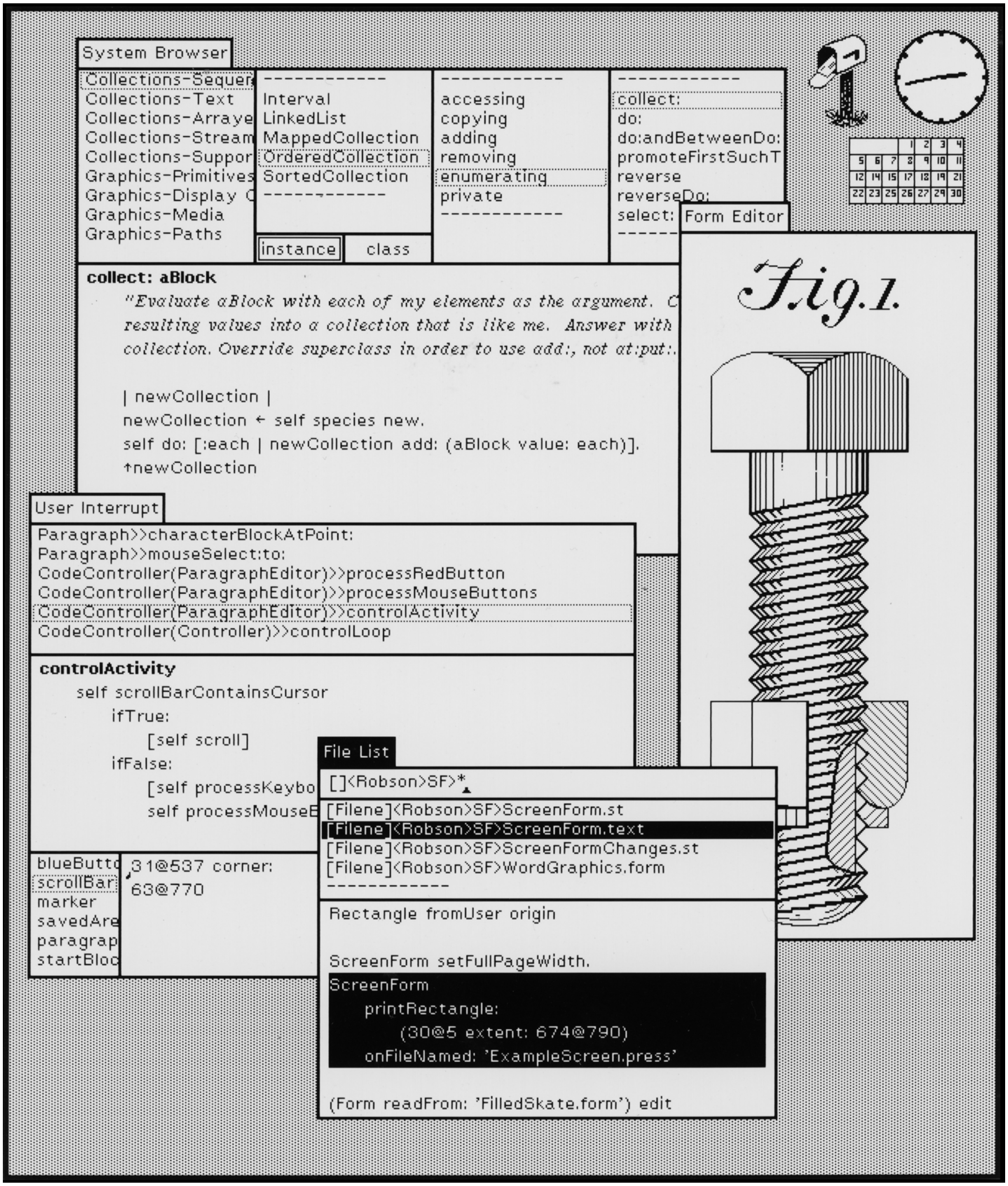



Figure 2.11: Smalltalk was a programming language and environment project in Xerox Palo Alto Research Center in the 1970s. The work of a handful of people, it became the direct antecedent for the modern graphical user interface.

A second implication of the task-artifact cycle is that continual exploration of new applications and application domains, new designs and design paradigms, new experiences, and new activities should remain highly prized in HCI. We may have the sense that we know where we are going today, but given the apparent rate of co-evolution in activity and artifacts, our effective look-ahead is probably less than we think. Moreover, since we are in effect constructing a future trajectory, and not just finding it, the cost of missteps is high. The co-evolution of activity and artifacts evidences strong hysteresis, that is to say, effects of past co-evolutionary adjustments persist far into the future. For example, many people struggle every day with operating systems and core productivity applications whose designs were evolutionary reactions to misanalyses from two or more decades ago. Of course, it is impossible to always be right with respect to values and criteria that will emerge and coalesce in the future, but we should at least be mindful that very consequential missteps are possible.



Figure 2.12: A-B: The Drift Table is an interactive coffee table; aerial views of England and Wales are displayed the porthole on top; placing and moving objects on the table causes the aerial imagery to scroll. This design is intended to provoke reaction and challenge thinking about domestic technologies.

The remedy is to consider many alternatives at every point in the progression. It is vitally important to have lots of work exploring possible experiences and activities, for example, on design and experience probes and prototypes. If we focus too strongly on the affordances of currently embodied technology we are too easily and uncritically accepting constraints that will limit contemporary HCI as well as all future trajectories.


Phone
Mail
Safari
Music
What can I help you with?

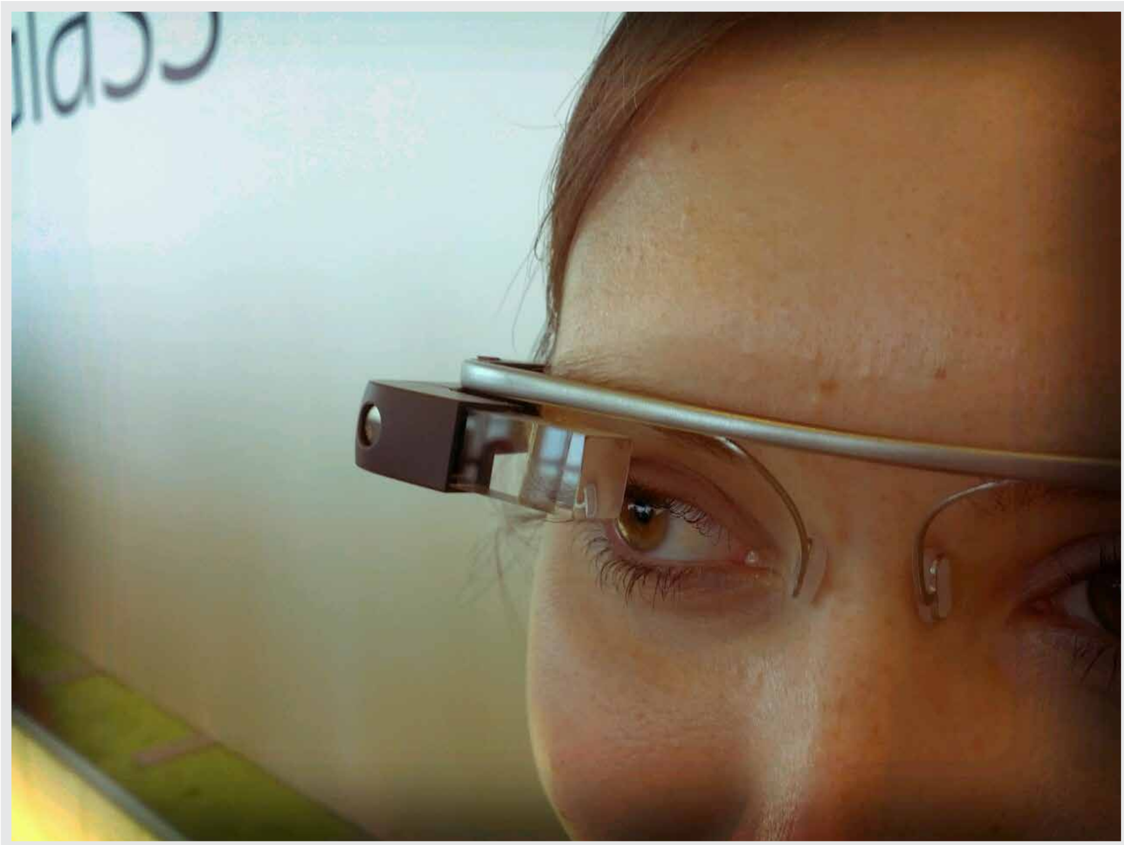


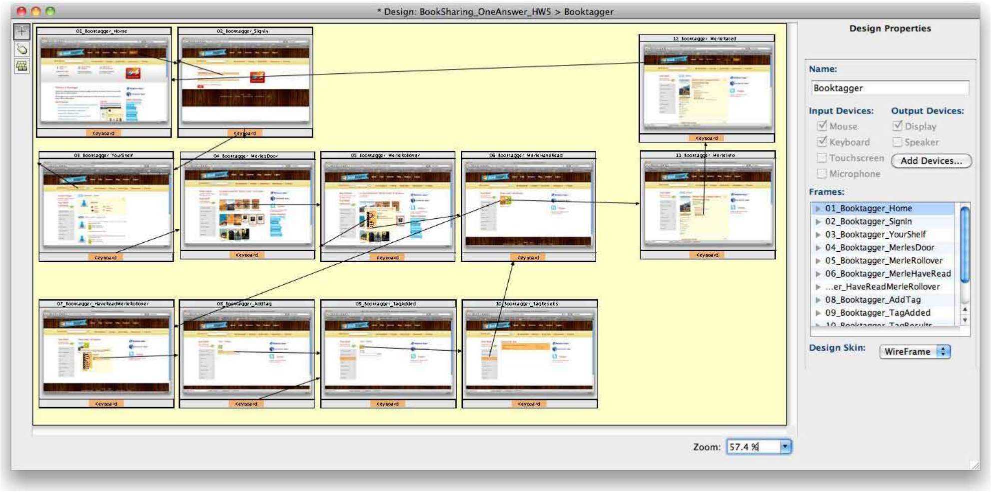



Figure 2.13 A-B: Siri, the speech-based intelligent assistant for Apple's iPhone, and the augmented reality glasses of Goggle's Project Glass are recent examples of technology visions being turned into everyday HCI experiences.

HCI is not fundamentally about the laws of nature. Rather, it manages innovation to ensure that human values and human priorities are advanced, and not diminished through new technology. This is what created HCI; this is what led HCI off the desktop; it will continue to lead HCI to new regions of technology-mediated human possibility. This is why usability is an open-ended concept, and can never be reduced to a fixed checklist.

## 2.5 A CALDRON OF THEORY

The contingent trajectory of HCI as a project in transforming human activity and experience through design has nonetheless remained closely integrated with the application and development of theory in the social and cognitive sciences. Even though, and to some extent because the technologies and human activities at issue in HCI are continually co-evolving, the domain has served as a laboratory and incubator for theory. The origin of HCI as an early case study in cognitive engineering had an imprinting effect on the character of the endeavor. From the very start, the models, theories and frameworks developed and used in HCI were pursued as contributions to science: HCI has enriched every theory it has appropriated. For example, the GOMS (Goals, Operations, Methods, Selection rules) model, the earliest native theory in HCI, was a more comprehensive cognitive model than had been attempted elsewhere in cognitive science and engineering; the model human processor included simple aspects of perception, attention, short-term memory operations, planning, and motor behavior in a single model. But GOMS was also a practical tool, articulating the dual criteria of scientific contribution *plus* engineering and design efficacy that has become the culture of theory and application in HCI.




*Copyright © Bonnie E. John. All Rights Reserved. Reproduced with permission. See section "Exceptions" in the copyright terms.*

FIGURE 2.14 A-B: CogTool analyzes demonstrations of user tasks to produce a model of the cognitive processes underlying task performance; from this model it predicts expert performance times for the tasks.

The focus of theory development and application has moved throughout the history of HCI, as the focus of the co-evolution of activities and artifacts has moved.?á Thus, the early information processing-based psychological theories, like GOMS, were employed to model the cognition and behavior of individuals interacting with keyboards, simple displays, and pointing devices. This initial conception of HCI theory was broadened as interactions became more varied and applications became richer. For example, perceptual theories were marshaled to explain how objects are recognized in a graphical display, mental model theories were appropriated to explain the role of concepts — like the messy desktop metaphor — in shaping interactions, active user theories were developed to explain how and why users learn and making sense of interactions. In each case, however, these elaborations were both scientific advances and bases for better tools and design practices.

This dialectic of theory and application has continued in HCI. It is easy to identify a dozen or so major currents of theory, which themselves can by grouped (roughly) into three eras: theories that view human-computer interaction as information processing, theories that view interaction as the initiative of agents pursuing projects, and theories that view interaction as socially and materially embedded in rich contexts. To some extent, the sequence of theories can be understood as a convergence of scientific opportunity and application need: Codifying and using relatively austere models made it clear what richer views of people and interaction could be articulated and what they could contribute; at the same time, personal devices became portals for interaction in the social and physical world, requiring richer theoretical frameworks for analysis and design.

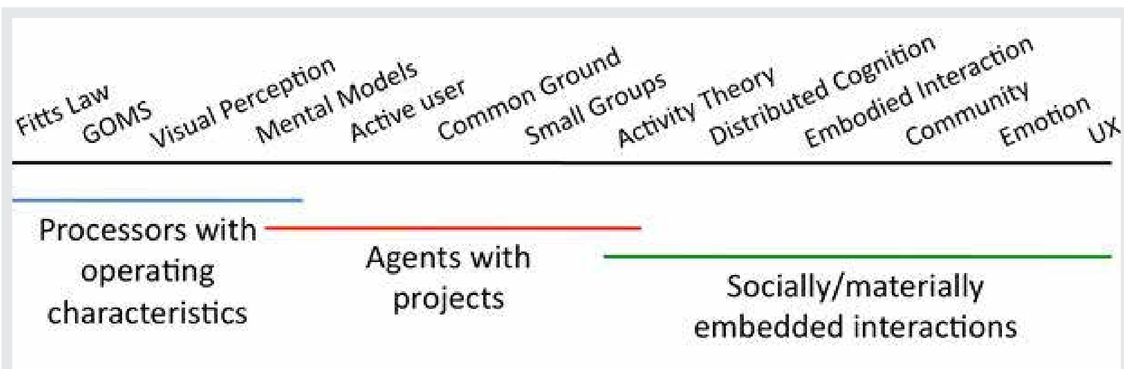



FIGURE 2.15: Through the past three decades, a series of theoretical paradigms emerged to address the expanding ambitions of HCI research, design, and product development. Successive theories both challenged and enriched prior conception of people and interaction. All of these theories are still relevant and still in use today in HCI.



The sequence of theories and eras is of course somewhat idealized. People still work on GOMS models; indeed, all of the major models, theories and frameworks that ever were employed in HCI are still in current use. Indeed, they continue to develop as the context of the field develops. GOMS today is more a niche model than a paradigm for HCI, but has recently been applied in research on smart phone designs and human-robot interactions.

The challenge of integrating, or at least better coordinating descriptive and explanatory science goals with prescriptive and constructive design goals is abiding in HCI. There are at least three ongoing directions — traditional application of ever-broader and deeper basic theories, development of local, sometimes domain dependent proto-theories within particular design domains, and the use of design rationale as a mediating level of description between basic science and design practice.

## 2.6 IMPLICATIONS OF HCI FOR SCIENCE, PRACTICE, AND EPISTEMOLOGY

One of the most significant achievements of HCI is its evolving model of the integration of research and practice. Initially this model was articulated as a reciprocal relation between cognitive science and cognitive engineering. Later, it ambitiously incorporated a diverse science foundation, notably social and organizational psychology, Activity Theory, distributed cognition, and sociology, and a ethnographic approaches human activity, including the activities of design and technology development and appropriation. Currently, the model is incorporating design practices and research across a broad spectrum, for example, theorizing user experience and ecological sustainability. In these developments, HCI provides a blueprint for a mutual relation between science and practice that is unprecedented.

Although HCI was always talked about as a design science or as pursuing guidance for designers, this was construed at first as a boundary, with HCI research and design as separate contributing areas of professional expertise. Throughout the 1990s, however, HCI directly assimilated, and eventually itself spawned, a series of design communities. At first, this was a merely ecumenical acceptance of methods and techniques laying those of beyond those of science and engineering. But this outreach impulse coincided with substantial advances in user interface technologies that shifted much of the potential proprietary value of user interfaces into graphical design and much richer ontologies of user experience.

Somewhat ironically, designers were welcomed into the HCI community just in time to help remake it as a design discipline. A large part of this transformation was the creation of design disciplines and issues that did not exist before. For example, user experience design and interaction design were not imported into HCI, but rather were among the first exports from HCI to the design world. Similarly, analysis of the productive tensions between creativity and rationale in design required a design field like HCI in which it is essential that designs have an internal logic, and can be systematically evaluated and maintained, yet at the same time provoke new experiences and insights.?á Design is currently the facet of HCI in most rapid flux. It seems likely that more new design proto-disciplines will emerge from HCI during the next decade.

No one can accuse HCI of resting on laurels. Conceptions of how underlying science informs and is informed by the worlds of practice and activity have evolved continually in HCI since its inception. Throughout the development of HCI, paradigm-changing scientific and epistemological revisions were deliberately embraced by a field that was, by any measure, succeeding intellectually and practically. The result has been an increasingly fragmented and complex field that has continued to succeed even more. This example contradicts the Kuhnian view of how intellectual projects develop through paradigms that are eventually overthrown. The continuing success of the HCI community in moving its meta-project forward thus has profound implications, not only for human-centered informatics, but for epistemology.

## 2.7 POINTERS: HOW TO LEARN MORE

In these "pointers" I have listed general background references to the discussion above, specific references to points made in the text, and reference to other chapters in the Encyclopedia of Human-Computer Interaction (Interaction-Design.org). I have organized the pointers by section, so the next six sections (below) echo the six major section headings in the paper itself (above).

### 2.7.1 Where HCI came from

There are many highly readable descriptions of the disciplinary landscape in which early HCI developed:

- 1980 volume of the journal Cognitive Science provides a vivid picture of the foundations of cognitive science as they were being built (http://csjarchive.cogsci.rpi.edu/1980v04/index.html);
- F. Brooks' book *The Mythical Man-Month* (1975, Addison-Wesley) is an insightful analysis of software engineering, and the original source for the idea that iterative prototyping is inevitable in the design and development of complex software;
- J. Foley and A. van Dam's book *Computer Graphics* (1982, Addison Wesley) describes the early field of computer graphics as a root of what would become human-computer interaction.

Vivid primary information about the founding of HCI - the proceedings of the 1982 US Bureau of Standards Conference in Gaithersburg, Maryland, are available in the ACM Digital Library at http://dl.acm.org/citation.cfm?id=800049

Several histories of HCI have been published:

- Carroll, J.M.?á (1997) Human-Computer Interaction: Psychology as a science of design. *Annual Review of Psychology, 48,* 61-83.?á (Co-published (slightly revised) in *International Journal of Human-Computer Studies, 46,* 501-522).
- Grudin, J. (2012) A Moving Target: The evolution of Human-computer Interaction. In J. Jacko (Ed.), *Human-computer interaction handbook: Fundamentals, evolving technologies, and emerging applications. (3rd edition)*. Taylor & Francis.
- Myers, B.A. (1998) A Brief History of Human Computer Interaction Technology. *ACM interactions*. Vol. 5, no. 2, March. pp. 44-54.

The leading HCI textbooks also include some discussion of history (see below).

### 2.7.2 From cabal to community

There is some dispute as to how to address the evolution of usability. In this overview, I take a historical view that the concept itself is evolving, analogous to way physics has treated its fundamental concepts, such as gravity and mass. See also

- Carroll, J.M. (2004) Beyond fun. *ACM interactions, 11(5),* 38-40.

The ACM Special Interest Group on Computer-Human Interaction (SIGCHI), and its CHI Conference, one of the most general and significant HCI conferences, now is explicitly organized into communities that manage pieces of the technical program (http://www.sigchi.org/communities). In fall of 2012, these communities included CCaA (Creativity, Cognition and Art), CSCW (Computer-Supported Cooperative Work), EICS (Engineering Interactive Computer Systems), HCI and Sustainability, HCI Education, HCI4D (HCI for Development), Heritage Matters, Latin American HCI, Pattern Languages and HCI, Research-practice Interaction, UbiComp (Ubiquitous Computing), and UIST (User Interface Software and Tools).

An even more diverse view of HCI can be appreciated by investigating HCI activities and interest groups embedded in professional communities *other than* ACM: the Design Research Society (designresearchsociety.org), the Association for Information Systems (sighci.org), the Human Factors and Ergonomics Society (hfes.org), the Society for Technical Communication (stc.org), the AIGA (aiga.org), International Communication Association (icahdq.org), the Interaction Design Association (http://www.ixda.org/), the IEEE Professional Communication Society (pcs.ieee.org), the European Association of Work and Organizational Psychology (eawop2013.org), and many others.

Further relevant material in the Encyclopedia of Human-Computer Interaction can be found in chapters 1, 3, 8, 13, 15, 19, 21, and 22.

### 2.7.3 Beyond the desktop

A classic discussion of the desktop metaphor is Apple Human Interface Guidelines: Apple & Raskin, J. (1992). *Macintosh Human Interface Guidelines*. Addison-Wesley Professional. ISBN 0-201-62216-5.

An early critique of the Macintosh user interface paradigm is:

- Gentner, D. and Nielsen, J. (1996) The Anti-Mac interface, *Communications of the ACM* **39**, 8 (August), 70-82.

The emergence of collaboration, mobility, and new types of user devices and interactions as major themes driving "HCI beyond the desktop" are discussed widely, of course; here are some starting points:

- Horn, D.B., Finholt, T.A., Birnholtz, J.P., Motwani, D. and Jayaraman, S. (2004) Six degrees of jonathan grudin: a social network analysis of the evolution and impact of CSCW research. In *Proceedings of the 2004 ACM conference on Computer supported cooperative work* (CSCW '04). ACM, New York, NY, USA, 582-591.
- Luff, P. and Heath, C. (1998) Mobility in collaboration. In *Proceedings of the 1998 ACM conference on Computer supported cooperative work* (CSCW '98). ACM, New York, NY, USA, 305-314.
- Shaer, O. and Hornecker, E. (2010) Tangible User Interfaces: Past, Present, and Future Directions. *Found. Trends Hum.-Comput. Interact*. 3, 1-2 (January), 1-137.
- Waller V. and Johnston, R.B. (2009) Making ubiquitous computing available. *Commun. ACM* 52, 10 (October 2009), 127-130.

Further relevant material in the Encyclopedia of Human-Computer Interaction can be found in chapters 4, 14, 23, and 27.

### 2.7.4 The task-artifact cycle

I use the term "task-artifact cycle" here, as originally introduced in a 1991 paper with Wendy Kellogg and Mary Beth Rosson, though think "activity" better conveys what I mean than "task". Not surprisingly, the terminology of the task-artifact cycle is itself an example of how HCI shifts under its own foundations; see

- Carroll, J.M., Kellogg, W.A., & Rosson, M.B. (1991) The task-artifact cycle.? In J.M. Carroll (Ed.),? *Designing Interaction: Psychology at the human-computer interface.* ? New York: Cambridge University Press, pages 74-102.

A good reference for the history of the Smalltalk project is

- Kay, A.C. (1996) The early history of Smalltalk. In *History of programming languages---II*, Thomas J. Bergin, Jr. and Richard G. Gibson, Jr. (Eds.). ACM, New York, NY, USA 511-598.

For more discussion of the drift table project, see

- Boucher, A. and Gaver, W. 2006. Developing the drift table. *interactions* 13, 1 (January 2006), 24-27.

For more discussion of the general point being emphasized through the example of the drift table, see

- Sengers, P. and W. Gaver, W. (2006) Staying open to interpretation: engaging multiple meanings in design and evaluation. In *Proceedings of the 6th conference on Designing Interactive systems* (DIS '06). ACM, New York, NY, USA, 99-108.

Further relevant material in the Encyclopedia of Human-Computer Interaction can be found in chapters 7, 12, 14, and 20.

### 2.7.5 A caldron of theory

I edited a book of theory overviews (Carroll, 2003), referenced below. It is available online (http://www.sciencedirect.com/science/book/9781558608085). I am currently curating theory overviews in the Synthesis Lectures on Human-Centered Informatics (http://www.morganclaypool.com/toc/hci/1/1).

- John, B.E. (2011) Using predictive human performance models to inspire and support UI design recommendations. In *Proceedings of the 2011 annual conference on Human factors in computing systems* (CHI '11). ACM, New York, NY, USA, 983-986.

Further relevant material in the Encyclopedia of Human-Computer Interaction can be found in chapters 5, 6, 9, 11, 16, 1 7, 24, 25, 26, and 28.

### 2.7.6 Implications of HCI for science practice and epistemology

The work I refer to on creativity and design rationale is collected in a book, Carroll (2012), reference below. A nice example, I think, of the theoretical multi-vocality I describe in this section can be appreciated by contrasting these three treatments of aesthetics in HCI design (all are published in the *Synthesis Lectures on Human-Centered Informatics*, http://www.morganclaypool.com/toc/hci/1/1):

- Hassenzahl, M. (2010). Experience Design: Technology for All the Right Reasons.
- Sutcliffe, A. (2009) Designing for User Engagement: Aesthetic and Attractive User Interfaces
- Wright, P. and McCarthy, J. (2010) Experience-Centered Design: Designers, Users, and Communities in Dialogue.

My reference to Kuhn regarding the development of science and knowledge is:

- Kuhn, T.S. (1962) The Structure of Scientific Revolutions. Chicago: University of Chicago Press.
- Kuhn, T.S. (1977) *The Essential Tension: Selected Studies in Scientific Tradition and Change*. Chicago and London: University of Chicago Press.


### 2.7.7 Textbooks

The number of important monographs is just too large to list, so I have concentrated in the list below on a few significant textbooks. Readers should also check the HCI Bibliography, the HCC Education Digital Library, the ACM Digital Library, and the Synthesis Series of lectures on human-centered informatics. These are the three most comprehensive textbooks:

- Dix, A.J., Finlay, J.E., Abowd, G.D. and Beale, R. (2003). *Human-Computer Interaction (3rd Edition)*. Prentice Hall
- Rogers, Y., Sharp, H. and Preece, J.J. (2011) *Interaction Design: Beyond Human-Computer Interaction (3rd ed.)*. John Wiley and Sons
- Shneiderman, B. and Plaisant, C. (2009). *Designing the User Interface: Strategies for Effective Human-Computer Interaction (5th ed.)*. Addison-Wesley

Several texts present more specialized views of HCI. Carroll (2003) collected a set of introductory papers on major theories used in HCI. L??wgren and Stolterman (2007) present a design perspective on HCI. Rosson and Carroll (2002) emphasize a software engineering view of HCI using a set of case studies to convey an engineering process view of usability. Tidwell (2011) presents a pattern-based approach to user interface design.

- Carroll, John M. (ed.) (2003). *HCI Models, Theories, and Frameworks: Toward a Multidisciplinary Science*. Morgan Kaufmann
- Löwgren, J. and Stolterman, E. (2007) *Thoughtful Interaction Design: A Design Perspective on Information Technology*. MIT Press.
- Rosson, M.B. and Carroll, J.M. (2002). *Usability Engineering: Scenario-Based Development of Human Computer Interaction*. Morgan Kaufmann.
- Tidwell, J. (2011) *Designing Interfaces (2nd ed.)*. O'Reilly Media.

### 2.7.8 Journals

The leading general journal for HCI is the ACM Transactions on Computer-Human Interaction. However, there are many other well-established journals of roughly equivalent quality: Human-Computer Interaction (emphasizes design research), Interacting With Computers, International Journal of Human-Computer Studies, *Behaviour and Information Technology*, International Journal of Human-Computer Interaction, Journal of Computer-Supported Cooperative Work. Recently, Association for Information Systems has initiated a Transactions on Human-Computer Interaction.Morgan-Claypool publishes a monograph series, Synthesis Lectures on Human-Centered Informatics.

My personal perspectives on the emergence and development of HCI are elaborated in several other articles, monographs, and introductions to edited books:

- Carroll, J.M. (1995) Introduction: The scenario perspective on system development. In Carroll, J.M. (Ed.), *Scenario-based design: Envisioning work and technology in system development*. New York: John Wiley & Sons, pp 1-17.
- Carroll, John M. (1997) Human-Computer Interaction: Psychology as a Science of Design. *Annual Review of Psychology, 48,* 61-

83.?á Co-published (slightly revised) in International Journal of Human-Computer Studies, *46(4), 501-522*

- Carroll, J.M. (1998) Reconstructing minimalism. In J.M. Carroll (Ed.) *Minimalism beyond?á "The Nurnberg Funnel"*. M.I.T. Press
- Carroll, J.M. (2000)?á *Making use: Scenario-based design of human-computer interactions.?á* MIT Press.?á Japanese edition published in 2003 by Kyoritsu Publishing; translated by Professor Kentaro Go.
- Carroll, John M. (2002). Human-Computer Interaction. In: (ed.), *MacMillan Encyclopedia of Cognitive Science*. Macmillan-Nature Publishing Group.
- Carroll, John M. (2004). Beyond fun. Interactions, *11(5), 38-40*
- Carroll, J.M. (2010) Narrating the Future: Scenarios and the Cult of Specification. In Selber, S. (Ed.), *Rhetorics And Technologies: New directions in writing and communication*. University of South Carolina Press, pp. 134-147.
- Carroll, J.M. (2010). Conceptualizing a possible discipline of Human-Computer Interaction. *Interacting with Computers, 22*, 3-12.
- Carroll, J.M. (2012) *The neighborhood in the Internet: Design research projects in community informatics*. Routledge.
- Carroll, J.M. (2012) Creativity and Rationale: The Essential Tension, in J.M. Carroll (Ed.) *Creativity and rationale: Enhancing human experience by design*. Springer, pages 1-10.

### 2.7.9 Relevant Conference Series

#### CHI - Human Factors in Computing Systems

| | | | | | | |
|---|---|---|---|---|---|---|
| 2011 | 2010 | 2009 | 2008 | 2007 | 2006 | 2005 |
| 2004 | 2003 | 2002 | 2001 | 2000 | 1999 | 1998 |
| 1997 | 1996 | 1995 | 1994 | 1993 | 1992 | 1991 |
| 1990 | 1989 | 1988 | 1987 | 1986 | 1985 | 1983 |
| 1982 | | | | | | |

#### ECSCW - European Conference on Computer Supported Cooperative Work

| | | | | | | |
|---|---|---|---|---|---|---|
| 2009 | 2007 | 2003 | 2003 | 2001 | 2001 | 1999 |
| 1997 | 1995 | 1993 | 1991 | 1989 | | |

Next conference is coming up 21 Sep 2013 in Paphos, Cyprus

#### CSCW - Conference On Computer-Supported Cooperative Work

| | | | | | | |
|---|---|---|---|---|---|---|
| 2012 | 2012 | 2012 | 2011 | 2010 | 2008 | 2006 |
| 2004 | 2004 | 2002 | 2000 | 1998 | 1996 | 1994 |
| 1992 | 1990 | 1988 | 1986 | | | |

Next conference is coming up 15 Feb 2014 in Baltimore, MD, USA

#### UIST - Symposium on User Interface Software and Technology

| | | | | | | |
|---|---|---|---|---|---|---|
| 2012 | 2012 | 2011 | 2010 | 2009 | 2008 | 2007 |
| 2007 | 2007 | 2006 | 2005 | 2004 | 2003 | 2003 |
| 2002 | 2001 | 2000 | 1999 | 1998 | 1997 | 1996 |
| 1995 | 1994 | 1993 | 1992 | 1991 | 1990 | 1989 |
| 1988 | | | | | | |

Next conference is coming up 08 Oct 2013 in St Andrews, UK

### NordiCHI - Nordic conference on human-computer interaction

2010 2008 2006 2004 2002 2002 2000
2000

Next conference is coming up 26 Oct 2014 in Helsinki, Finland

### BCSHCI People and Computers

2012 2010 2009 2008 2006 2006 2005
2004 2003 2002 2001 2000 1998 1997
1996 1995 1994 1993 1992 1991 1989
1988 1987 1986 1985

Next conference is coming up 09 Sep 2013 in London, UK

### SIGGROUP - Conference on Supporting Group Work

2010 2009 2007 2005 2003 2001 1999
1997 1995 1993 1991 1990 1988 1986
1984 1982

### DIS - Designing Interactive Systems

2012 2010 2008 2006 2004 2002 2000
1997 1995

### CC - Creativity and Cognition

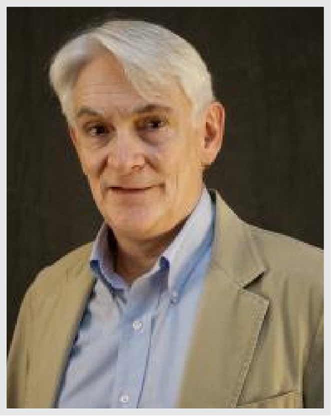
2011 2009 2007 2005 2002 1999

## 2.8 REFERENCES

Myers, Brad A. (1998): *A Brief History of Human-Computer Interaction Technology*. In Interactions, 5 (2) pp. 44-54

## ABOUT THE AUTHOR

### John M. Carroll





HAS ALSO PUBLISHED UNDER THE NAME OF:
"John M Carroll", "John Carroll", "J. M. Carroll", and "J. M. Carroll"

PERSONAL HOMEPAGE:
http://jcarroll.ist.psu.edu/

CURRENT PLACE OF EMPLOYMENT:
The Pennsylvania State University

John M. Carroll is Edward Frymoyer Chair Professor of Information Sciences and Technology at the Pennsylvania State University. His research interests include methods and theory in human-computer interaction, particularly as applied to networking tools for collaborative learning and problem solving, and the design of interactive information systems. His books include Making Use (MIT Press, 2000), HCI in the New Millennium (Addison-Wesley, 2001), Usability Engineering

(Morgan-Kaufmann, 2002, with M.B. Rosson) and HCI Models, Theories, and Frameworks (Morgan-Kaufmann, 2003), Rationale-based software engineering (Springer, 2008, with J. Burge, R. McCall and I. Mistrik), and Learning in Communities (Springer, 2009). He serves on several editorial boards for journals, handbooks, and series and is Editor-in-Chief of the ACM Transactions on Computer-Human Interactions. He received the Rigo Award and the CHI Lifetime Achievement Award from ACM, the Silver Core Award from IFIP, the Alfred N. Goldsmith Award from IEEE. He is a fellow of the ACM, IEEE, and the Human Factors and Ergonomics Society.

## YOUR NOTES AND THOUGHTS ON CHAPTER 2

Record your notes and thoughts on this chapter. If you want to share these thoughts with others online, go to the bottom of the page at:
http://www.interaction-design.org/encyclopedia/human_computer_interaction_hci.html

**NOTES:**

# CHAPTER 3

# User Experience and Experience Design

*by Marc Hassenzahl.*

I open my eyes. Lush light floods the room, birds chatter. It is only 6:30 o'clock in the morning, but I feel well-rested and alive; time to get up, to brew some coffee. Are you jealous of my morning routine? Were you startled out of your sleep by a merciless alarm clock? Was it dark outside, no birds around, and did you feel groggy and bleary-eyed?

This chapter is about experiences created and shaped through technology (aka *User Experience*) and how to deliberately design those. The wake-up experience created by an alarm clock substantially differs from the experience created by sunrise and happy birds. The question is whether we can create technology which understands the crucial features of sunrise and birds and which succeeds in delivering a similar experience, even when the sun refuses to shine and the birds have already left for Africa.

In fact, the experience I described in the beginning was not created by sun and birds, but by Philips' *Wake-Up Light*. This is a crossing of an alarm clock and a bedside lamp. Half an hour before the set alarm, the lamp starts to brighten gradually, simulating sunrise. It reaches its maximum at the set wake-up time and then the electronic birds kick in to make sure that you really get up. Admittedly, it is a surrogate experience, but so are love stories and travel novels. It is artificial, but not vulgar. And more importantly, it substantially changes the way one wakes up. It changes the experience. The object itself, its form, is rather unremarkable (see Figure 3.1).

Figure 3.1: Philips' Wake-Up Light.

The Philips *Wake-Up Light* has nevertheless the power to "transcend its encasing" because its contribution is not one to the aesthetics of things, but to the aesthetics of experiences. This is the challenge designers and vendors of interactive products face: Experience or User Experience is not about good industrial design, multi-touch, or fancy interfaces. It is about transcending the material. It is about creating an experience through a device.

Video 3.1: Marc's introduction to User Experience and Experience Design.

Video 3.2: Marc's advice on designing with experience in mind.

*Courtesy of Rikke Friis Dam and Mads Soegaard. Copyright: CC-Att-ND (Creative Commons Attribution-NoDerivs 3.0 Unported). View full screen or download (800 KB). View transcription/captions.*

Video 3.3: Marc's main guidelines and ethical considerations.

*Courtesy of Rikke Friis Dam and Mads Soegaard. Copyright: CC-Att-ND (Creative Commons Attribution-NoDerivs 3.0 Unported). View full screen or download (1 MB). View transcription/captions.*

Video 3.4: Future directions.

*Courtesy of Rikke Friis Dam and Mads Soegaard. Copyright: CC-Att-ND (Creative Commons Attribution-NoDerivs 3.0 Unported). View full screen or download (800 KB). View transcription/captions.*

I will start this chapter with a discussion of our Western societies' shift from the material to the experiential and the potential problems technology-oriented businesses have in accommodating this shift. *User Experience* and *Experience Design* can be a remedy to this by bringing experience to the fore. I then discuss *Experience* and *User Experience* to flesh out a view which has the potential to advance the way we will design future technologies. I end with some examples of *Experience Design* and finally offer a simple model of *Why*, *What* and *How* as a starting point for the enthusiastic *Experience Designer*.

## 3.1 FROM THE MATERIAL TO THE EXPERIENTIAL

In Roald Dahl's *Charlie and the Chocolate Factory,* young Charlie faces a tough choice. He just found the last *Golden Ticket* in a bar of *Whipple-Scrumptious Fudgemallow Delight*. (Figure 3.2). It is one of only five invitations to visit Willy

Wonka's legendary chocolate factory. Charlie is promised a day full of "mystic and marvellous surprises that will entrance, delight, intrigue, astonish, and perplex beyond measure. In your wildest dreams you could not imagine that such things happen to you!"

Figure 3.2: Charlie Bucket discovers his *Golden Ticket*.



But Charlie is poor. It is a freezing winter and the whole family of seven is living on not more than cabbagy meals and the occasional boiled potato. People already offered as much as $500 for the ticket. Wouldn't it be more sensible to forfeit Wonka's frivolous offer and to secure the money? In the end, Charlie took the ticket and was awarded with the most extraordinary experience of his life.

Charlie chose the experience over the material. He could have had a winter coat or fire wood instead of the experience, but he already knew that only the visit to the chocolate factory has the power to add some meaning to his life. In fact, studies show that *experiential* purchases (i.e., the acquisition of an event to live through, such as a concert, a dinner, a journey) make people more happy than *material* purchases (i.e., the acquisition of tangible objects, such as clothing, jewellery, stereo equipment) of the same value (Boven and Gilovich 2003; Carter and Gilovich 2010).

In a series of studies, Leaf van Boven and colleagues (2010) further uncovered stigmatizing stereotypes: Participants characterized people with a material orientation as self-centred, insecure, or judgmental, but people with an experiential orientation as humorous, friendly, open-minded, intelligent, caring, or outgoing. The seemingly negative stance towards the materialistic is an indication of a *post-materialistic* culture. Ronald Inglehart (1997) argued that societies in sustained periods of material wealth become increasingly interested in values such as personal improvement. They transform into highly individual *Experience Societies* (Schulze 1992; Schulze 2005) whose members equate happiness with the acquisition of positive life events. Decried as superficial and consumerist in the 80ties and 90ties of the last century, we now witness a version of the *Experience Society* which favours meaningful engagement to earning money and begins to dissociate experience and expenditure. Experiences are no longer supposed to be available at exotic places only. They can be close by: a day out in the sun, working the garden, a barbecue with friends, or a trip to the local flea market. In the foreword to the 2005 edition of his book, Gerhard Schulze (2005, p IX) mentions some signifiers of the new millennium's *Experience Society*: deceleration instead of acceleration, less instead of more, uniqueness instead of standardisation, concentration instead of diversion, and making instead of consuming. All these are not necessarily associated with material wealth. Admittedly, to develop a post-materialistic (i.e., experiential) orientation may require sufficient food, clothing, and shelter (Inglehart 1997; Maslow 1954). This is the gist of Charlie Bucket's dilemma: choosing a frivolous one-day experience in a chocolate factory over supporting his family with food and clothing seems almost immoral. However, while I agree that an experiential orientation in life requires some food, clothes and shelter as a necessary precondition (Inglehart 1997), I do not believe that it needs caviar, Gucci, and a chateau in the hills of the Cote d'Azur. Most of us in the developed countries have the basis for leading a post-materialistic life.

## 3.2 EXPERIENCE AND BUSINESS

Though the transformation to a post-materialistic experience society has been recognized by business, as indicated by books such as *The Experience Economy* (Pine & Gilmore, 1999) or *Experiential Marketing* (Schmitt 1999), it still struggles with making sense of it. A good example is the music industry. While the number of concerts is still rising, record sales dropped considerably from 2000 and onwards. For example, Madonna's *Confessions on a Dancefloor* sold only 1.6 million copies, but her world tour generated about 200 million dollars. According to *Pollstar* (Bongiovann 2010), in 2009 the average ticket price for a top 100 act in the US was about $64, a CD made only $13.99. Typically, illegal digital downloads are made responsible for this effect. But the missing willingness to pay for music in the form of a tangible product may also be a consequence of shifting from a materialistic to an experiential orientation. Today the music itself matters, not ownership, argues Arthur Schock (2010), a booker for independent electronic artists, in a recent interview for *Spiegel-Online*. He reported on record release parties with 800 raving guests but only ten records sold afterwards. "Liking bears no relation to buying the CD," he concluded. On *Creative Deconstruction* Rich Huxley (2010) mused: "If we can all now make, distribute and sell music, to succeed we've got to differentiate ourselves from the crowd and give people something they can't get elsewhere. If we can give people something that isn't repeatable and isn't copyable then all the better. So, what's unique and not copyable? A feeling, or an experience." Instead of complaining about declining CD sales, the music industry must develop new, more experiential formats.

Why aren't they? One of the main reasons why the music industry dislikes the shift away from the material is the limited scalability of experience. Once produced, a CD can be copied and sold in theoretically infinite quantities, while an artist can only play a limited number of concerts a year, with a limited number of paying attendants. As long as most industries and their strategies are still geared towards earning money by mass-producing and selling tangible objects, their take on the experiential is often not more than a feeble marketing strategy. For example, the *German Telekom* recently made “experiencing” its marketing claim (“Erleben, was verbindet”). The companion website promises to be a place for sharing memorable and unique experiences. But a close look reveals hardly more than the occasional sponsored live event interspersed with badly disguised attempts to sell standard products and services. Experience is considered a vehicle for marketing, but not understood as the very product that is sold. The transition from an economy of products and services to one of experience and transformation certainly requires more (Pine and Gilmore 1999). This is the challenge we face: *Experience* or *User Experience* is not about technology, industrial design, or interfaces. It is about creating a meaningful experience through a device.

## 3.3 THE EVASIVE BEAST CALLED USER EXPERIENCE

Experience is an almost overwhelmingly rich concept, with a long history of debate and many attempts to “define” it (Jay 2004). I primarily focus on experiences as meaningful, personally encountered events (in German: “Erlebnis”) and not so much on the knowledge gained through these events (in German: “Erfahrung”). These experiences are *memorized* stories of use and consumption and distinct from the *immediate* moment-by-moment experience (e.g., Forlizzi and Battarbee 2004; Kahneman 1999). While the immediate moment-by-moment experience is certainly interesting, memorized experience is of more practical relevance. This is simply because most of our waking time, we are feasting on vivid memories of the past (or anticipations) rather than on immediate pleasures.

The construction of experiences as stories from moment-by-moment experience is not straightforward. For example, experiences tend to improve over time. As van Boven (2005, p. 137) puts it: “As one forgets the incidental annoyances and distractions that detract from the online, momentary enjoyment of an experience, one’s memory of an experience can be sharpened, levelled, and ‘spun’ so that the experience seems better in retrospect than it actually was.” Who doesn’t fall victim to a “rosy view” now and then. We are spinning - not necessarily consciously - our own experiences.

But what is *in* an experience? Psychologically, an experience emerges from the integration of perception, action, motivation, and cognition into an inseparable, meaningful whole. The intimate relation between those single concepts is reflected by, for example, Russell's (2003) model of emotions, which stresses the importance of cognitive processes, such as self-observation, attribution, and categorization, for the experience of emotions. And most action theories (e.g., Kaptelinin & Nardi, 2006; Carver & Scheier, 1989) assume close links between actions, thoughts and emotions. In sum, I argue for understanding experience as "an episode, a chunk of time that one went through [...] sights and sounds, feelings and thoughts, motives and actions [...] closely knitted together, stored in memory, labelled, relived and communicated to others. An experience is a story, emerging from the dialogue of a person with her or his world through action" (Hassenzahl 2010, pp. 8). An experience is subjective, holistic, situated, dynamic, and worthwhile.

While an experience is a complex fabric of feelings, thoughts, and actions, I believe emotions and fulfilment of universal psychological needs to have an accentuated role. Although emotions are certainly complex, they all share an inherent evaluation, pleasure and pain, which provide "the yardstick on which qualitatively different possibilities can be compared" (Russell 2003, p. 153). This evaluation is important in shaping future behaviour and - if positive - one source of happiness. But where does the pleasure come from? Sheldon and others (2001) demonstrated the intimate link between the pleasantness of an experience and the fulfilment of universal psychological needs in that experience, such as the need for autonomy or stimulation. A specific look at technology-mediated positive experiences revealed stimulation, relatedness, competence, and popularity as the salient sources of pleasure (Hassenzahl et al. 2010). Being asked for a recent positive experience with technology, a young woman provided the following example: "I was on a short trip to Dublin. In the early hours, my mobile phone woke me up. My boyfriend, who stayed at home, had just texted a sweet 'I love you' (Figure 3.3). This is an example of a relatedness experience, which gets its positive meaning through fulfilling a need for social relationship and intimacy.

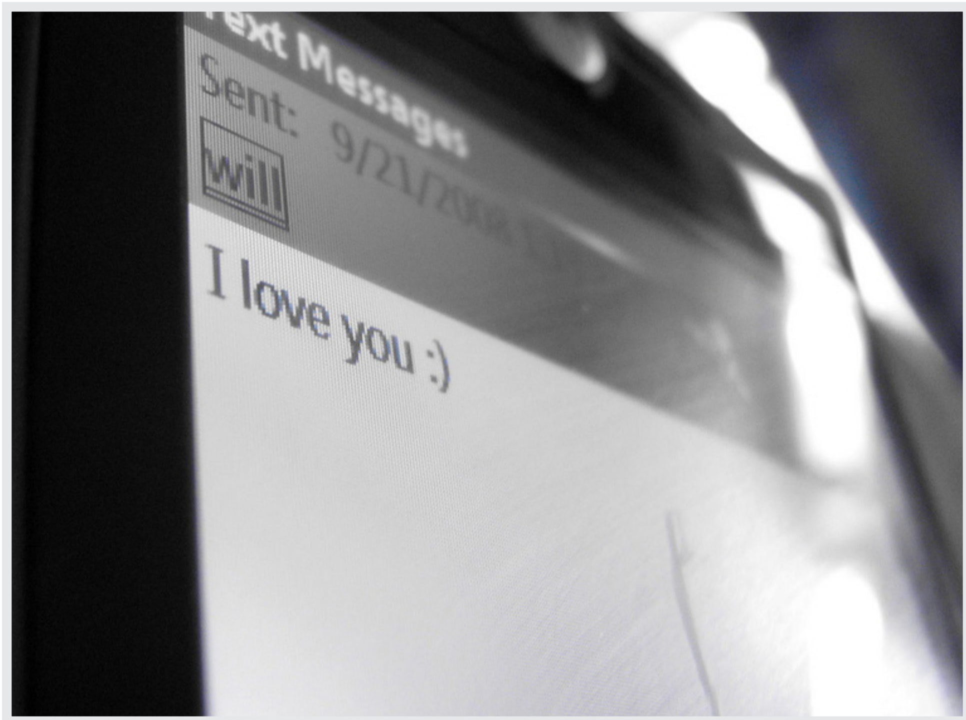



Figure 3.3: A sweet 'I love you'



The mobile phone is instrumental to creating this experience, but the positive emotions and the meaning are evoked through the fulfilment of a universal psychological need. Need-fulfilment is what makes an experience pleasurable.

Usage and consumption always translate into an experience, a story of use, a story of consumption: just like a rollercoaster becomes embedded into a (hopefully) meaningful, emotion-laden story of a rollercoaster ride, full-blown with stimulation, excitement and enjoyment. Seemingly different products and situations are represented in a similar format - that of experience. Thus, as long as we focus on the experiences created and shaped by interactive products, we may not distinguish *User Experience* from *Experience* in general. User Experience is just a sub-category of experience, focusing on a particular mediator - namely interactive products. If it comes to actual *Experience Design*, that is the question of how to deliberately create and shape experiences, a distinction between interactive products and other mediators of experiences may be helpful, but does not seem crucial.

The perspective on *Experience* and *User Experience* developed here should not be understood as definite. It is a starting point for debate, an attempt to advance a concept of *Experience* and *User Experience* that will change the way interactive products are - hopefully to the better.

## 3.4 EXPERIENCE DESIGN: DESIGNING THE POST-MATERIALISTIC

With the sharp distinction between the experiential and the material suggested by many authors (e.g., Boven and Gilovich 2003), an "experiential interactive product" appears like a contradiction in terms. While experience is intangible, volatile, an interactive product is tangible, a mass-produced piece of technology. The "electronic gadget" is the very prototype of a material purchase. The seasoned post-materialist, though, ceases to strive for yet another novel communication device. She will rather enjoy writing a letter (Figure 3.4).


Figure 3.4: Letters to an English schoolgirl.



But even the post-materialist's experience is most of the time mediated. Writing a letter requires a pen, paper, and a messenger, who in turn needs a carriage, a zeppelin, or a plane. This holds for all the typical examples of experiential purchases provided by van Boven (2003): travel requires transportation, dining requires a good kitchen, and a concert requires instruments and amplification. Things are not the opposite of experiences, but create and substantially shape them. The combination of a pen and a piece of paper, and the resulting activity of writing with one's own hand, has certain features which in turn shape the resulting experience. It is, for example, relatively slow and, thus, offers time for reflection, not provided by more efficient technologies (Lindley et al 2009). Thus, the post-materialist is not necessarily a "green luddite" (Kozinets 2007) who shuns technology in general. But she is more interested in the experience created than taking pride in the ownership of the product or technology that created it. Once created, the experience is what is owned - an immaterial, personal story. The product is only of interest as it is identified as being crucial in creating the experience (Hassenzahl et al 2010).

The challenge of designing interactive products for the post-materialist is to bring the resulting experience to the fore - to design the experience before the product. Or as Buxton (2007, p. 127) puts it: "Despite the technocratic and materialistic bias of our [US-American] culture, it is ultimately experiences we are designing, not things." But what does that mean, *to design an experience?* For Buxton it seems a matter of how it feels to act through a product, in the moment it is used - the moment-by-moment experience. He used different orange squeezers to highlight how different usage can "feel" even if the function remains the same. This addresses the *How* of product use, the *Aesthetics of Interaction*. This notion of *Experience* - as focusing on *how* something is done - was notably sparked by the success of Apple's iPhone, featuring a so far unique aesthetic of interaction, but basically fulfilling the same tasks as any other mobile phone.

While certainly important, reducing experience to the mere "pleasure due to the feel of the action" (Buxton 2007, p. 129) is not doing justice to its multifaceted nature. Conceptually, the broad view of *Experience* as meaningful stories has much more to offer than a narrow view as pleasurable, moment-by-moment feeling. Take the story of the young woman on a trip to Dublin from the preceding section. The experience gets its positive feel and meaning through the fulfilment of a need for relatedness, a need for feeling close to relevant others. The story speaks of intimacy, expressed, for example, by the liberty to send the message very early in the morning. The man was confident that his girlfriend would not be annoyed by the message. And while receiving love messages is always a wonderful thing, being in a foreign place, far away from home, certainly intensified this experience. In this example, the mobile phone was used as a tool for creating a relatedness experience. But the mobile phone is neither especially adapted to this, nor does it in any way imply the creation of this experience. It is nothing more than an awkward piece of infrastructure: even with the most elegant shell or navigational structure, it does not reflect the love put into the message. To give another example: While a telephone is certainly able to connect distant lovers, it embodies a strictly conversational model. However, feeling close is not about good conversations only, it is a about a feeling of presence and emotional expression. The telephone is not exceptionally good at this - as Peter Robinson observed in *All the Colours of Darkness* after a late night telephone conversation between Inspector Banks and his Sophia:

"'Goodnight' said Banks. And the last thing he heard was her laughter as she puts down the phone. Banks felt more alone and further away for having just talked to Sophia than he had before the call. But it was always like that - the telephone might bring you together for a few moments, but there's nothing like it for emphasising distance."

We have all experienced the awkward silence when we have run out of stories to tell while not wanting to hang up on our loved one. This is the result of a misfit between the conversational model embodied by a telephone and the psychological requirements of a relatedness experience.

This must not necessarily be so, as the prevalent research on technology-mediated intimacy demonstrates (e.g., Vetere et al 2005). An unpublished review (Heidecker et al 2010) counted 144 published concepts of alternative communication devices, most of them much better adapted to the requirements of "feeling close" than any commercially available mobile phone. In many cases, the technological innovation embedded in those novel devices is negligible - they neither feature elaborated new algorithms nor future materials or fancy interface concepts. Their superiority is due to the intimate understanding of certain experiences, feelings, situations, boundary conditions, and how those experiences can be created and shaped through a thing.

The post-materialistic interactive product is, thus, not so much a tangible object, but a story transported or told through an object - a "material tale" or "psychosocial narrative". Dunne (2006, p. 69) explains: "[... B]ehavior is a narrative experience arising from the interaction between our desire to act through products and the social and behavioural limitations imposed [...] through [their] conceptual models." We will inevitably act through products, a story will be told, but the product itself creates and shapes it. The designer becomes an "author" creating rather than representing experiences.

So far, there are not many commercially available products, which reflect the notion of *Experience Design* as the creation of meaningful stories through a product. An exception is FM3's Buddha Machine, dubbed the *Anti-iPod* by the *Wall Street Journal* in 2007 (Wagstaff 2007) - see Figure 3.5.

Figure 3.5: The Buddha Machine II.



The *Buddha Machine* is an electronic device loaded with nine ambient loops (Version 2.0) produced by FM3, an experimental music duo from China. It plays back one of those loops in 8-bit quality through an inbuilt speaker, has a button to skip through the loops, a knob to change the pitch of the playback and its volume. That's it. The *Buddha Machine* is a meditative experience. It tells a story of contemplation rather than restless consumption and suggests a way of doing so.

In 2007, the *Buddha Machine* was an unexpected commercial success with over 50.000 units sold. Joshua wrote on *Resident Advisor*: "With the Buddha Machine, FM3 have unwittingly unleashed a real phenomenon: [...] a personal stereo, a musical toy, a Buddhist souvenir, and a conceptual commodity offering valuable lessons for our consumption-obsessed times." He quotes Christiaan Vi-rant, one of the creators of the *Buddha Machine*: "That's the beauty of the Buddha Machine, it's really … serendipity." And Rob Walker (2007) noted for the *New York Times Magazine*:

> "*And of course there's the anti-iPod factor: the relief of not having to make a choice in a world awash with entertainment and self-expression options. Moreover, at a moment when the unused abilities of feature-loaded computers, cellphones and even microwave ovens pile up faster than we can keep track of them, it's satisfying to know that once you've turned the Buddha Machine on, you are using it to its full capacity.*"

The *Buddha Machine* is an example of a device, which "manages to transcend the cheap plastic frame in which it's encased" (Heater 2008). It is a technology that offers a meaningful, valuable, and aesthetic experience and not just a bunch of functions, leaving it to the users to figure out how to incorporate them into their daily lives.

The *Buddha Machine* is an excellent example of a full-blown post-materialistic device. However, one may easily view it a representative of a novel product genre, which coexists with more "practical" genres, but does not affect the design of those more practical products. I disagree. A post-materialistic, experiential orientation can potentially be loaded into every product. An example is Swantje Krauß' diploma design project, which I supervised together with Olaf Barski. Krauß set out to design a new type of improved "bucket" for the grape harvest. Typically, grapes are picked by hand, gathered in a bucket, which is then emptied into a larger container. This bucket is clearly a tool; its design a tough exercise in practicality and classical ergonomics. However, Krauß added an interesting feature beyond the obvious: The bucket can be transformed into a seat (see Figure 3.6) which allows the vintager to take a rest from her physically demanding work.

Figure 3.6: From a tool to a place to rest.

This seemingly small detail is interesting for at least three reasons. First, the bucket embeds both activities - gathering grapes and taking a rest - on an equal level, making clear that a rest is accepted as an integral part of the overall activity. Second, the bucket has to be empty to be transformed into a seat. This reflects upon the admittedly puritan ideal of “business before pleasure” and functions as a clear signal for the “appropriate” moment for taking a rest. Krauß’ design makes sure that its users either pick grapes or rest, but resting and still doing a little bit of cleaning or sorting the grapes is impossible. This implies a clear separation between work and rest - an important psychological requirement for having a truly re-creative break. Third, the bucket suggests a particular way of taking that rest, namely *in* the vineyard, contemplating and enjoying the views or having a chat with colleagues.

Admittedly, a bucket is not a typical exemplar of an interactive product. Nevertheless, Krauß’ example shows that understanding grape picking as more than a mere task, as an experience packed with psychological needs, emotions and meaning enables the designer to become an author of stories conveyed through the product.

## 3.5 WHY, WHAT AND HOW

Let me summarize my thoughts on *Experience* in a simple conceptual model. I distinguish three different levels, when designing an experience through the interaction with an object: The *Why*, *What* and *How* level.

The *What* addresses the things people can do through an interactive product, such as “making a telephone call,” “buying a book,” or “listening to a song.” Reflected by a products’ functionality, the *What* is often intimately tied to the technology itself or a certain product genre. The *How* in turn addresses acting through an object on an operational, sensory-motor level: Buttons pressed, knobs turned, menus navigated, touch screens stroked, or remotes waggled. The *How* is even more tied to the actual object to be designed and its context of use.

The *How* is the typical realm of the interaction designer: to make given functionality accessible in an aesthetically pleasing way. To give an example: "Making a telephone call" (a *What*) requires an action to select a conversional partner, as well as to initiate and to end the call. *How* this is done with - let's say - a mobile phone is specified by the interaction designer. The example of the different orange squeezers, Bill Buxton (2007) provided, addresses possible differences in the quality of the interaction design, the *How*. Even given the same functionality (i.e., squeezing oranges), performing the action "feels" better with some products. Nowadays, the bundle of *What* and *How* is typically considered the product, and an especially sensual, aesthetic, novel, or stimulating arrangement of interaction makes this product "experiential."

This view ignores peoples' actual motivation to use a product. For the couple being separated, the SMS was not primarily an SMS, it was a love message, a way to fulfil their need for relatedness. This is the *Why* of product use. Telephone calls are not only - technologically speaking - telephone calls. In reality, they are the glorious beginning or the sad end of a close relationship, a surrogate good-night kiss, an act of support, a way to kill time, or a pizza order. People engage in these activities out of a need to be related, to help, to be stimulated, or to ease their appetites. The telephone just happens to be instrumental, but it does not necessarily reflect upon the underlying needs, emotions, and associated practices.

*Experience Design* is a remedy to this. It starts from the *Why*, tries to clarify the needs and emotions involved in an activity, the meaning, the experience. Only then, it determines functionality that is able to provide the experience (the *What*) and an appropriate way of putting the functionality to action (the *How*). Experience Design wants the *Why*, *What* and *How* to chime together, but with the *Why*, the needs and emotions, setting the tone (see Figure 3.7). This leads to products which are sensitive to the particularities of human experience. It leads to products able to tell enjoyable stories through their use or consumption.

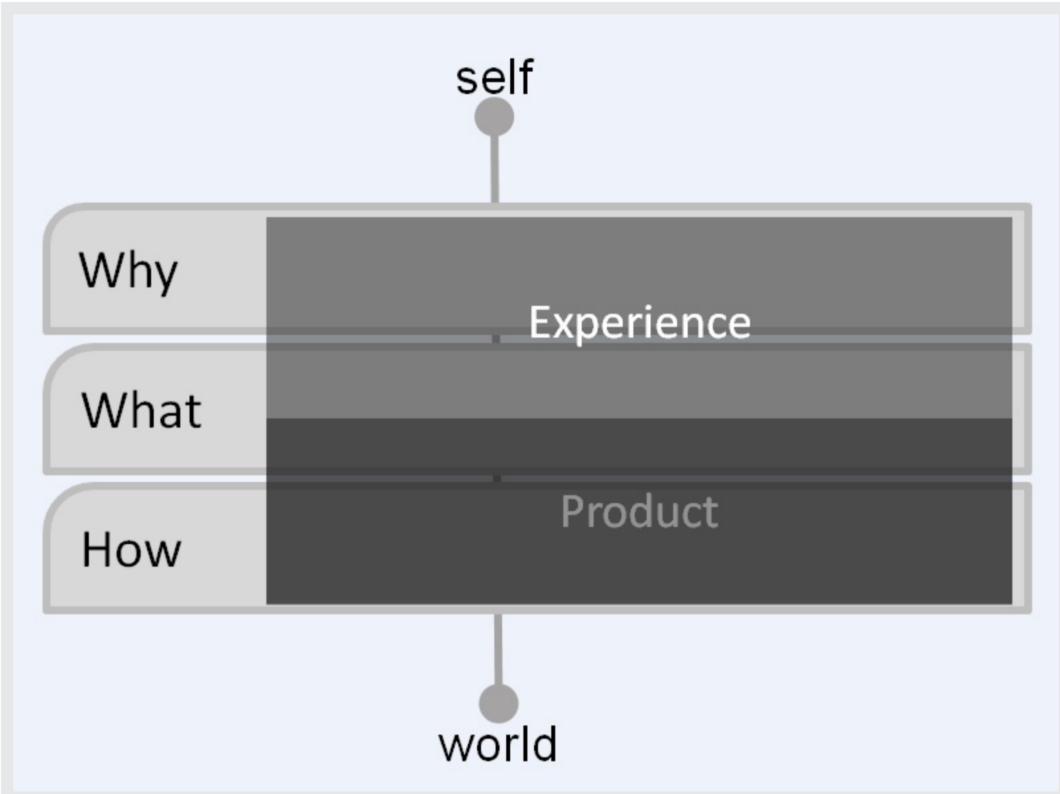



FIGURE 3.7: From the *Why* to the *What* and the *How*: Three levels to consider when designing technology-mediated experiences.



## 3.6 CONCLUSION AND FUTURE DIRECTIONS

The notion of *(User) Experience* as stories told through products has a potential to change the way we think and design. At the moment, the majority of commercially available interactive devices is either too practical or too open-ended. The practical view results in very obvious and uninspiring stories: how exciting is keeping a calendar on a mobile phone? The open-ended view on the other hand just provides functionality, such as texting, and leaves it to the user to come up with meaningful and inspiring usage scenarios, such as sending "love messages." In this case, the creation of meaningful experiences through appropriating a technology remains the responsibility of the "user". In contrast, *Experience Design* stands for technology, which suggests meaningful, engaging, valuable, and aesthetically pleasing experiences in itself. Thinking "communication experiences" rather than "mobile devices" opens up a huge design space for possible devices - even slippers (Chen et al 2006) or pillows (Laschke et al 2010).

Don't get me wrong, we still need all the wonderful technologies, dreamt up by engineers and computer scientists all over the world. But they are only materials - canvas, colours, and brushes - for the *Experience Designer*. From a business perspective, shifting attention from technological to experiential advancement makes sense, as long as the invention of new technologies and their marketing becomes increasingly difficult. Just take 3D television as an example: It is an innovation born out of a frantic need for re-inventing television to ensure future sales. The result is an expensive, hard to sell technology, without much power to impact our lives "The new movie by Darren Aronofsky now in 3D! So what?" Indeed, other technology-mediated innovations, such as improving the social experience of watching television as a family or over a distance, require less effort in terms of resources (both on the vendor and the consumer end), but at the same time offer a profound improvement of current practices and according experiences. We should definitely shift attention (and resources) from the development of new technologies to the conscious design of resulting experiences, from technology-driven innovations to human-driven innovations.

## 3.8 COMMENTARY BY ERIC L. REISS

How to cite this commentary in your report

### Eric L. Reiss

© *Eric L. Reiss*

Eric Reiss has been meddling with multimedia and web projects for longer than he cares to remember. Born in San Antonio, Texas, in 1954, Eric Reiss has held a wide range of eclectic jobs including ragtime piano player (in a St. Louis house of ill-repute), senior copywriter (in an ad-house of ill-repute). In November, 2000, his book, Practical Information Architecture was published by ...

Eric L. Reiss
Eric L. Reiss is a member of The Interaction Design Foundation

For years, I have struggled to understand the difference between "user experience" and "experience". I couldn't help but smile as Marc struggled with this same problem. In fact, by the penultimate paragraph, Marc had decided to place the word "user" in parenthesis. This supports the viewpoint that both Marc and I seem to share, that both of these terms mean essentially the same thing, despite the semantic bickering in the professional community. Listen to the first few minutes of Marc's first video for some very succinct remarks on this matter.

Of course, if one really feels a need to differentiate between "user experience" and "experience", Marc has some interesting comments and observations. In the introduction, he suggests that "It is about creating an experience through a device." "It" is the elusive beast in the current debate.

Later on, Marc states, "Experience or User Experience is not about technology, industrial design, or interfaces. It is about creating a meaningful experience through a device." I agree 100% with the first statement, but I question the second part; I don't think that experience is necessarily related to a device. Certainly, Charlie's experience with the chocolate factory didn't involve "experience through a device" unless you pedantically define the golden ticket as a "device". (The presence of a device or lack thereof often lies at the heart of the "user experience vs. experience" debate.)

But let's take things a step further. If I go out to greet the sunrise - not courtesy of Philips, but standing in my garden on a glorious spring day - my experience does not depend on technology, industrial design, or interfaces. Since I like sunrises, my limbic system is busy distributing dopamine - a reward chemical that affects my mood. And my body is soaking up Vitamin D, which improves my health. There are no devices involved in this interaction between me and the sun (accompanied by soft dew on the grass between my toes, birds chirping, and that undefinable smell released by vegetation as it, too, awakes and greets the sun).

As a designer, I see user experience (UX) as the perception left in someone's mind following a series of interactions between people, devices, and events - or any combination thereof. "Series" is the operative word.

Some interactions are active - clicking a button on a website, giving a waiter your order at a restaurant, getting out of the rain at a picnic.

Some interactions are passive - viewing a beautiful sunrise will trigger the release of reward chemicals in our brain. This applies to any and all of our five senses.

Some interactions are secondary to the ultimate experience - the food tastes good because the chef chose quality ingredients and prepared them well. The ingredients are good quality because the farmer tended his fields. The crop interacted well with the rain that year.

Of course, all interactions are open to subjective interpretation - some people don't like celery or sunrises. Remember, a perception is always true in the mind of the perceiver; if you think sunrises are depressing, there's little I can say or do to convince you otherwise. However, this is why designers often fall back on "best practice" - most people react favorably to sunrises.

For these reasons, I think that designing a "user experience," represents the conscious act of:

- coordinating interactions that are controllable (choosing food ingredients, training waiters, designing and programming buttons)
- acknowledging interactions that are beyond our control (uncomfortable seats in a 100-year-old theater, lack of fresh produce in winter, low-hanging clouds that hide the sky)
- reducing negative interactions (providing tents as emergency shelters at outdoor events in case of rain; making sure restaurant seating next to the noisy kitchen door is the last to be filled, putting in an extra intermission so folks can stretch their legs)

A good user-experience designer needs to be able to see both the forest and the trees. That means user experience has implications that go far beyond usability, visual design, and physical affordances. As UX designers, we orchestrate a complex series of interactions and the emotional responses and/or physical responses that these interactions generate. To look at "experience" in terms of individual service or product touchpoints is ultimately too limiting. It is the total sum of that counts.

Another interesting point is contained in Marc's example of the "I love you" SMS. Here, the phone's designer merely facilitated an interaction between two individuals. Facilitating an experience and creating one are two very different things - designers should always consider which role they being asked to play at any given time in the design process.

Finally, the value of an experience is exceptionally subjective. I was delighted to see Marc's reference to the van Boven and Gilovich work from 2003. This ties in directly to the work of Akerlof, Spence, and Stiglitz on asymmetric information, which won them the Nobel Prize in Economics in 2001. Let me share some thoughts.

Despite any theoretical shift from a materialistic society that covets things, to a post-materialistic culture that nurtures experience, the value of physical items has always increased if they are accompanied by a good story. A vintage watch is worth more if it comes with all its original paperwork, receipts, etc. An antique chair's value can change dramatically depending on its provenance. (A chair previously owned by Winston Churchill is going to be worth more than a chair from my house). Yet neither watches nor wing chairs physically change because they come with a piece of paper.

As designers, dealing with the subjective nature of experience could well be our greatest challenge. This may also explain why experience is so difficult to define - which brings us back full circle to the beginning of this commentary.

## 3.9 COMMENTARY BY DONALD A. NORMAN

How to cite this commentary in your report

### Donald A. Norman

*© Donald A. Norman*

Don Norman is the author of numerous books including “Emotional Design,” and more recently, “Living with Complexity.” He is co-founder of the Nielsen Norman group, a professor at KAIST (in Korea), and IDEO fellow, and a design theorist, studying the fundamentals of modern design.Donald A. Norman has a Bachelor of Science in Electrical Engineering from MIT and a Doctor of Philosophy in Ps...

Donald A. Norman
Donald A. Norman is a member of The Interaction Design Foundation

Technologies migrate as they mature. In early childhood, their very existence is a marvel, even as people wonder what can be made of it. In early adolescence, they become more and more able to perform useful functions for us, and for a while, they are judged primarily on their ability to do more and more, better and better. Finally, in maturity, it is the quality of the experience provided by these technologies that matter. Adolescents thrust their technological underpinnings into our consciousness, even as we resisted. But once the technology becomes mature, it recedes into the background, supportive of the total experience it provides.

Design, it has been said (Krippendorff, 1989) is the creation of meaning, and as Hassenzahl points out, the essence of meaning to us people is our experiences.

The chapter by Marc Hassenzahl ought to be required reading in courses of design, and perhaps even more importantly, in engineering and computer science. Do the devices we design and produce work well? Do they do marvelous, mysterious operations, working invisibly across space and time? Yes, they do, but doing that is a means, not the end. The end is the experiences they engender, the stories we tell, and the way that they enriched our lives.

But this creates a problem. We know how to design things that accomplish particular, concrete actions. But how can experiences be designed? As Hassenzahl points out they can't be: they can only be supported. To use another design term: we can design in the *affordances* of experiences, but in the end it is up to the people who use our products to have the experiences.

The product provides the "How" part of an experience. It is up to people to provide the "What" and the "Why." But designers can help here as well, setting the framework, providing the initiative, providing examples.

Design has moved from its origins of making things look attractive (styling), to making things that fulfill true needs in an effective understandable way (design studies and interactive design) to the enabling of experiences (experience design). Each step is more difficult than the one before each requires and builds upon what was learned before.

The first step toward experiences was to learn about and embrace emotion and products that were pleasurable. This step has just been taken, with an increasing number of books, journal articles, and conferences attesting to the interest in this topic. But these steps too were in their infancy, addressing primarily the need and desirability along with the technical difficulties of measuring the resulting emotions or pleasure.

Marc Hassenzahl throws down the gauntlet for future designs: to produce products that deliver the Why, What, and How.

### References

Krippendorff, Klaus (1989). On the Essential Contexts of Artifacts or on the Proposition That "Design Is Making Sense (Of Things)". Design Issues, 5(2), pp. 9-39.

### About the Author

Don Norman is the author of numerous books including "Emotional Design," and more recently, "Living with Complexity." He is co-founder of the Nielsen Norman group, a professor at KAIST (in Korea), and IDEO fellow, and a design theorist, studying the fundamentals of modern design. Although he invented the term "User experience" while an executive at Apple, he is pleased that people like Marc Hassenzhal have moved beyond the phrase to deep substance. Although Norman travels an inordinate amount, he can always be found at www.jnd.org

## 3.10 COMMENTARY BY MARK BLYTHE

How to cite this commentary in your report

### Mark Blythe

© *Mark Blythe*

Mark is Professor of Interdisciplinary Design at Northumbria University. He is a design ethnographer working in the field of Human Computer Interaction. His research interests include the ageing population, looming ecological catastrophe and the messed up world we're in. Recent projects include collaborations with Goldsmiths' Interaction Research Studio which were based at a residential ...

Mark Blythe
Mark Blythe is a member of The Interaction Design Foundation

## The Hitch Hiker's Guide to Experience Design

I like the idea of Marc's alarm clock, but I suspect that with the right sort of hangover, the gentle birdsong might sound like horses galloping over a tin bridge. It is an old point now but perhaps worth repeating: no experience can be guaranteed. In McCarthy and Wright's formulation experiences cannot be designed they can only be designed for (3). It is sometimes countered that, on the contrary, experiences get designed all the time and we only have to look at film, theatre and the other arts to see how. But in some ways there couldn't be a worse example. Consider James Cameron's "Avatar". Incredibly, the person who I was with found it... boring, yes, boring, if you can believe it. As I sat in open mouthed astonishment at the technological and artistic achievement, my friend's jaw dropped only to yawn. Experience cannot be guaranteed even with Hollywood budgets. But conceding that you can't please all of the people all of the time does not necessarily mean that we cannot learn anything from understandings of literature, film and other media. Novelists, dramatists and film directors have, after all, been designing impossible things for a very long time.

Somewhere in the fourth video of the very interesting conversations with Marc he points out that the differences in how we tell stories matter. The way that we tell and understand story is crucial for experience design and for this reason there is as the saying goes "a small but growing body of work" that draws on critical theory (e.g. 1.2.3). Critical theory developed from the study of literature, drama and film. It is a catch all term that covers a very diverse range of perspectives such as psycho-analysis, feminism and deconstruction. Because its subject of study is the everyday - novels, films, TV, it is often dismissed as pretentious. It draws on specialized or high falutin' terms like "dialogical" and can seem unnecessarily abstract and difficult.

The terms "dialogical" and "monological" were key to the thought of the literary critic Mikhail Bakhtin and they are increasingly applied to experience design (e.g. 3). Such concepts can be quite confusing and are perhaps best explained with reference to the kinds of contexts Bakhtin thought about. For Bakhtin a "monological" utterance expects no answer. For instance, the order "Charge!" on a battlefield anticipates action, not debate. For Bakhtin the style of narrative in the novels of Tolstoy is also monological. The narrator of Anna Karenina for example knows the most intimate thoughts and actions of every character. How he knows what everyone is thinking is not at issue. In Dostoevsky on the other hand the narrative is dialogical. In Notes from the Underground the narrator constantly tries to anticipate and guess the reader's responses to what he is saying. "*You imagine no doubt, gentlemen, that I want to amuse you. You are mistaken in that, too. I am by no means such a mirthful person as you imagine*." Even where a neutral narrative voice begins the novel with omnipresent authority, as in the Brothers Karamazov, it is later revealed that this voice belongs to one of the characters with a partial perspective and sources which might or might not be reliable. Today monological authority is increasingly undermined by dialogical forms. To find examples today we might look at the anti Mubarak protests in Egypt in 2011, organized in part through Facebook and Twitter.

Unlike a paper based encyclopedia resources such as this are also dialogical in the form of commentaries and invitations for reader responses. Douglas Adams predicted this long ago when he wrote that the Hitch Hiker's Guide to the Galaxy had supplanted the older and more pedestrian Encyclopedia Galactica in many of the more relaxed civilizations on the Outer Eastern Rim of the Galaxy. This, he said, was partly because it was slightly cheaper but mostly because it had the words “Don’t Panic:” inscribed in large friendly letters on the front cover. The guide’s field workers like Ford Prefect, would travel the universe, write about it and send their copy to the editors. Towards the end of his life Adams became less interested in writing novels and more involved in developing new media such as the pioneering game Starship Titanic. The HG2G website which attempted to create a real hitch hiker’s guide with user generated content was a precursor to wikipedia. What then might a hitch hiker’s guide to experience design look like?

“*Experience design is complicated, really complicated. I mean, you may think planning a holiday in Centre Parks is needlessly difficult, but that’s just peanuts to experience design*”

The style might settle down after this and go on to note that it draws on many other disciplines - psychology, sociology and yes even literary theory sometimes. Many years ago, as another new field of study formed, Roland Barthes pointed out that interdisciplinary work is not achieved by gathering a number of sciences around a new “subject”: “Interdisciplinarity” he said “consists in creating a new object that belongs to no one”. Or as Marc Hassenzahl puts it in conversation here, the many relevant disciplines must talk to each other to find what works.

## References

1. Bardzell J & Bardzell S (2008) Interaction criticism: a proposal and framework for a new discipline of hci Interaction Design CHI ‘08 extended abstracts on Human factors in computing systems
2. Blythe, M. and Cairns, P. 2009. Critical methods and user generated content: the iPhone on YouTube. CHI ‘09. (Boston, MA, USA, April 04 - 09, 2009

3. McCarthy J., and Wright P., (2008) Technology as Experience. MIT Press Cambridge Mass.

## 3.11 COMMENTARY BY WHITNEY HESS

How to cite this commentary in your report

### Whitney Hess

© ***Whitney Hess***

Whitney Hess is an independent user experience designer based in New York City. She helps make stuff easy and pleasurable to use. Whitney is a senior experience design consultant with Happy Cog, an advisor to RedStamp, and consults with a variety of startups and major corporations. She was the UX consultant on Boxee's beta app, and conducted extensive user research for the United Stat...

Whitney Hess
Whitney Hess is a member of The Interaction Design Foundation

Experience designers make a career out of alleviating people's problems and bringing them joy. To do this we remove barriers to entry; reduce the gap between current knowledge and target knowledge; make information easy to find and share; present content in digestible, understandable formats; adapt to the context of use; and above all else, maximize the potential for people to succeed.

But before we can find the right solution (the "How"), we need to carefully prioritize the features we invest in (the "What"). To determine our priorities, we need to clearly and accurately define the problem (the "Why"). To determine the problem, we need to identify the intended audience (the "Who"). And to determine the audience, we need to listen. At its core, this is the goal of experience design.

The result of a well-defined *who*, *why*, and *what* is an elegant *how* that can provide people with new dimensions of understanding, productivity, and pleasure.

Products, services and systems should improve the quality of people's lives, reduce stress, and create effeciencies that didn't previously exist. But uninformed design often yields the wrong *what* for the wrong *who* with a poorly considered *how*, causing pain and confusion, adding unnecessary cost, and defeating the purpose of the entire effort.

Marc Hassenzahl superbly demonstrates the power of a triumphant *how*, by doing what the best experience designers do: wrapping us up in a story and taking us where we need to go. His opening to this chapter illustrates the vastness of an experience designer's purview - what I am only able to convey with arms stretched wide - by making us feel what we would have felt had we been experiencing the products for ourselves.

Like with writing, just about all digital experiences are a proxy for meatspace, and that is a hurdle we must constantly overcome. Hassenzahl helps us recognize that not only do we have the responsibility to create the closest approximation of the mental, emotional and spiritual experience, but we also have the opportunity to use technology to create fulfilling experiences that would never be materially, logistically, or viably possible in the physical space.

"Experience or User Experience is not about good industrial design, multi-touch, or fancy interfaces," Hassenzahl states. "It is about transcending the material. It is about creating an experience through a device." I have chosen to read his use of "device" in its perhaps secondary but broader meaning - "a plan, scheme, or trick with a particular aim," rather than as a material object. I don't view experience design as contingent upon having an object (digital or material) with which to interact, though I can understand the argument of being device-centric given the implications of there being a *user*. But a user of a service might engage solely through conversation, with the designed "device" being *flow*. An experience designer's ultimate output is a plan itself, with all its conditions and contingencies. The success of the experience lies in the thoroughness and thoughtfulness of the plan.

Hassenzahl's guidelines on how to craft experiences as a dialog between designer and participant are insightful and well-articulated. He finally finds the common ground amongst what many practitioners have been preaching for decades: our work is a matter of ethics. We have to live with the choices we make, so let's do whatever it takes to be proud of the stories we tell and experiences we create.

I would like to see experience designers spend less time preoccupied with technology and more time exploring new environments on the hunt for new experiences. If we are bound to our chairs and screens, how will we discover new ways to create connection, meaning and purpose, and promote those learnings to the people we serve - or better yet, discover for ourselves whom we wish to serve in the first place?

## 3.12 COMMENTARY BY PAUL HEKKERT

How to cite this commentary in your report

### Paul Hekkert

© *Paul Hekkert*

Paul Hekkert is professor of Form Theory at the department of Industrial Design of Delft University of Technology. His main research interest is product experience, including product aesthetics, emotion, expressiveness, and attachment. Next, he is involved in design methodology and has co-developed an interaction-centred design approach, called ViP (Vision in Product design)....

Paul Hekkert
Paul Hekkert is a member of The Interaction Design Foundation

Everything I read in Marc Hassenzahl's chapter sounds so true, so valuable, and so familiar. It comforted me, and it puzzled me. Of course, designers are there to shape experiences. Nokia, or any mobile phone manufacturer, is all about connectedness, not about these mobile devices. All design starts with a 'why', and next comes the 'what', or the 'how'. I will come back to this order in a minute.

Recently, I published a new book together with designer and colleague Matthijs van Dijk: “Vision in Design: A guidebook for innovators” (2011). It spends over 300 pages on explaining (future) designers on how to find the WHY of their designs, what we call its ‘raison d’être’ or ‘Daseinsgrund’. The reason of existence is grounded in the future world, as the designer sees it, and reflects what the designer wants to offer people given this world. What do you want people to understand, see, be able to, feel or experience? This goal or ultimate reason is indeed often experiential[11]. A random example from a student: “I want passengers to experience a sense of freedom within the limited space of an aircraft, by stimulating mental travel.” This experiential goal takes into account the context of an aircraft, the mental state people are in during traveling, social anxiety, and people’s love of mind wandering. This ultimate experience comes first; the product is (just) instrumental in realizing it. Or as Marc puts it: it’s all about bringing “… the resulting experience to the fore - to design the experience before the product.” Hence the familiarity.

It is so obvious and logical and yet Marc has managed to phrase it in a way that is crisp and clear and thereby he opens it up to an audience that may not seem so aware of its logic. People who are caught up in technological advancements for their own sake? And here Marc also puzzles me. Why this emphasis on interactive products? All products are – in essence – interactive, they allow for and require interaction, and all products can contribute to, stage, shape, facilitate, or enable experiences. Think of Starck’s Juicy Salif, designed to stimulate conversations between a son- and mother-in-law. Or the example of the bucket “… a bucket is not a typical exemplar of an interactive product”, as Marc writes. So what?

Somewhere along the line, the ‘why’ must be transformed in a ‘what’ and a ‘how’ and I believe the crux is in changing the order Marc proposed. After the WHY, designers should not immediately follow with a WHAT, but first decide

1. When we look at public products for instance, the underlying reason can also be behavioral. Designing for behavioral change is currently a popular topic in design research.

on the HOW. This HOW is the user experience or product experience as we have coined it (Schifferstein and Hekkert, 2008); it captures the way people will interact with and experience the to-be-designed product that is not yet defined. These qualities of the interaction are intangible and not bound to anything, and they determine whether the ultimate experiential goal will be met. The WHAT that is next to be designed is simply a carrier of these qualities and it can be any type of product, an interactive product, a service, or a web application.

Let us consider an example of how this works. Years ago, student Sanne Kistemaker defined the experiential goal (the WHY) of her design project as "I want people to experience sarcastic triumph while staring at other people". We all want this, right? Watch other people while on the train, see what habits they have composed, see how they interact each other. She could have easily solved her design goal by some augmented reality type of application on a smart phone. But before she decided on the 'carrier' of her experience, she first defined what the interaction with the product should be. This interaction, as she saw it, should be tricky, reluctant, apprehensive, and straightforward, to make the user feel guilty and hesitating, yet proud and rebellious. She *designed* this interaction and corresponding user experience without having a clue what product should do this. And she beautifully solved it without any technological means: a newspaper that is offered with a small hole in the middle (see Figure)[2]. You may hesitate to use this paper, it is a little tricky, but once you do, in a train or on a terrace, you will certainly feel guilty and probably experience a sarcastic triumph!

2. Itis crucial for the design that the whole is already there; the experience would fail if the user had to make his/her own hole.


metro
WK Special
Eindexamens goed gemaakt
Britney Spears is emotioneel wrak
Online pesten verergert 'gewoon' pesten
ViP 2006 - sanne kistemaker - 1055798




actually you do WANT to watch...
but you know you can't!

To sum up, Marc importantly stresses that the user experience, the experience of the (interactive) product, should never be an end in itself, but is always instrumental to some life experience. All very true. And since the final design, the product, is again instrumental to the user experience, it seems only logical to make this the order of things: ultimate experience (WHY) > user experience (HOW) > product (WHAT). Designing along these lines is exactly what Verganti (2009) means when he speaks of 'design-driven innovation', where the designer *pushes* a new meaning, a new experience onto the public. Marc's chapter gives us many clues on how to do this.

## References

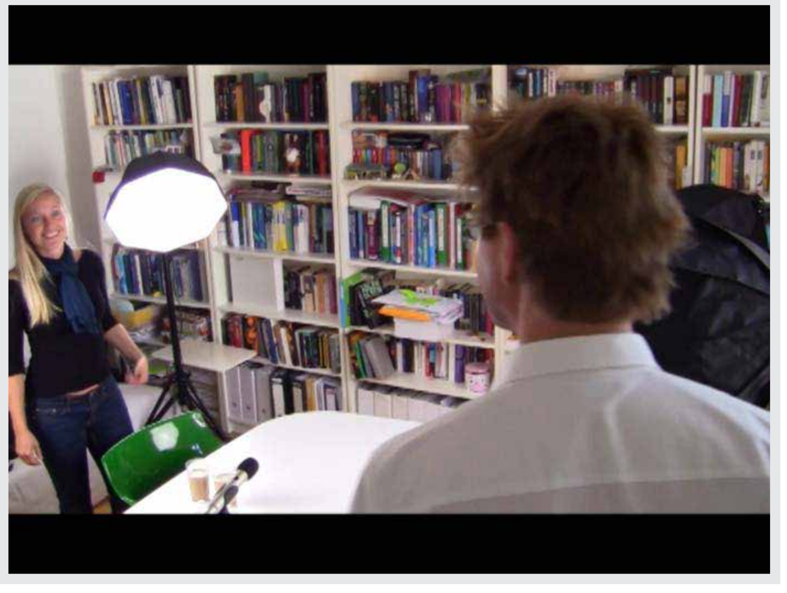
Hekkert, P. & van Dijk, M. (2011). Vision in Design: A Guidebook for Innovators. Amsterdam: BIS.

Schifferstein, H.N.J. & Hekkert, P. (Eds.)(2008). Product Experience. Amsterdam: Elsevier.

Verganti, R. (2009). Design Driven Innovation – Changing the Rules of Competition by Radically Innovating what Things Mean. Boston, MA: Harvard Business Press.

## 3.7 BEHIND THE SCENES

Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.



Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.


*Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.*


*Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.*



Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.

Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.

*Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.*

## 3.13 REFERENCES


*Bongiovann, Gary (2010). PollStar 2010 Midyear Business Analysis. Retrieved 9 November 2010 from PollStar Pro: http://www.pollstarpro.com/specialfeatures2010/2010MidYearBusinessAnalysis.PDF*

*Boven, Leaf Van (2005): Experientialism, Materialism, and the Pursuit of Happiness. In Review of General Psychology, 9 (2) pp. 132-142*

*Boven, Leaf Van and Gilovich, Thomas (2003): To Do or to Have? That Is the Question. In Journal of Personality and Social Psychology, 85 (6) p. 1193–1202*

*Boven, Leaf Van, Campbell, Margaret C. and Gilovich, Thomas (2010): Stigmatizing Materialism: On Stereotypes and Impressions of Materialistic and Experiential Pursuits. In Personality and Social Psychology Bulletin, 30 (4) pp. 551-563*

*Buxton, Bill (2007): Sketching User Experiences: Getting the Design Right and the Right Design. Morgan Kaufmann*

*Carter, Travis J. and Gilovich, Thomas D. (2010): The Relative Relativity of Material and Experiential Purchases. In Journal of Personality and Social Psychology, 98 pp. 146-159*

*Carver, Charles S. and Scheier, Michael F. (2001): On the Self-Regulation of Behavior. Cambridge University Press*

*Chen, Chun Y., Forlizzi, Jodi and Jennings, Pamela (2006): ComSlipper: An expressive* *design to support awareness and availability. In: Proceedings of the CHI 06 Conference on Human Factors in Computing Systems - extended abstracts on Human factors in computing systems 2006. pp. 369-374*

*Dunne, Anthony (2006): Hertzian Tales: Electronic Products, Aesthetic Experience, and Critical Design. MIT Press*

*Forlizzi, Jodi and Battarbee, Katja (2004): Understanding experience in interactive systems. In: Proceedings of DIS04: Designing Interactive Systems: Processes, Practices, Methods, & Techniques 2004. pp. 261-268*

*Hassenzahl, Marc (2010): Experience Design: Technology for All the Right Reasons. Morgan and Claypool Publishers*

*Hassenzahl, Marc, Diefenbach, Sarah and Göritz, Anja (2010): Needs, affect, and interactive products - Facets of user experience. In Interacting with Computers, 22 (5) pp. 353-362*

*Heater, Brian (2008). Hands On: Buddha Machine 2. Retrieved 9 November 2010 from Gearlog: http://www.gearlog.com/2008/11/hands_on_buddha_machine_2.php*

*Heidecker, S., Hassenzahl, M., Eckoldt, K., & Hillmann, U. (2010). Designing for relatedness experiences: A Review of technologies mediating intimate relationships. in preparation*

*Huxley, Rich (2010). Music is Not Our Currency. Retrieved 9 November 2010 from Creative Deconstruction: http://www.creativedeconstruction.com/2010/05/music-is-not-our-currency/*

*Inglehart, Ronald (1997): Modernization and Postmodernization: Cultural, Economic, and Political Change in 43 Societies. Princeton University Press*

*Jay, Martin (2006): Songs of Experience: Modern American and European Variations on a Universal Theme. University of California Press*

*Kahnemann, Daniel (2003): Objective happiness. In: Kahneman, Daniel, Diener, Ed and Schwarz, Norbert (eds.). "Well-Being: The Foundations of Hedonic Psychology". Russell Sage Foundation Publicationspp. 3-25*

*Kaptelinin, Victor and Nardi, Bonnie A. (2007): Acting with technology: Activity theory and interaction design. In First Monday, 12 (4)*

*Kozinets, Robert V. (2008): Technology/Ideology: How Ideological Fields Influence Consumers' Technology Narratives. In Journal of Consumer Research, 34 (6) pp. 864-881*

*Laschke, Matthias, Hassenzahl, Marc and Mehnert, Kurt (2010): linked.: a relatedness experience for boys. In: Proceedings of the Sixth Nordic Conference on Human-Computer Interaction 2010. pp. 839-844*

*Lindley, Sian E., Harper, Richard and Sellen, Abigail (2009): Desiring to be in touch in a changing communications landscape: attitudes of older adults. In: Proceedings of ACM CHI 2009 Conference on Human Factors in Computing Systems 2009. pp. 1693-1702*

*Maslow, Abraham H. (1954): Motivation and Personality. HarperCollins Publishers*

*Meggitt, Joshua (2007). FM3: Made in Beijing. Retrieved 9 November 2010 from Residentadvisor.net: http://www.residentadvisor.net/feature.aspx?836*

*Pine, B. Joseph (1999): The Experience Economy: Work Is Theater & Every Business a Stage. Harvard Business Press*

*Russell, James A. (2003): Core Affect and the Psychological Construction of Emotion. In Psychological Review, 110 (1) p. 145–172*

*Schmitt, Bernd H. (1999): Experiential Marketing: How to Get Customers to Sense, Feel, Think, Act, Relate. Free Press*

*Schock, Artur (2010). Gut finden und kaufen steht in keinem Verhältnis. Retrieved 9 November 2010 from Spiegel Online: http://www.spiegel.de/netzwelt/web/0,1518,710269,00.html*

*Schulze, Gerhard (2005): Die Erlebnisgesellschaft: Kultursoziologie der Gegenwart.*

*Schulze, Gerhard (1992): Die Erlebnisgesellschaft: Kultursoziologie der Gegenwart. Campus Verlag*

*Sheldon, Kennon M., Elliot, Andrew J., Kim, Youngmee and Kasser, Tim (2001): What Is Satisfying About Satisfying Events? Testing 10 Candidate Psychological Needs. In Journal of Personality and Social Psychology, 80 (2) pp. 325-339*

*Vetere, Frank, Gibbs, Martin R., Kjeldskov, Jesper, Howard, Steve, Mueller, Florian, Pedell, Sonja, Mecoles, Karen and Bunyan, Marcus (2005): Mediating intimacy: designing technologies to support strong-tie relationships. In: Proceedings of ACM CHI 2005 Conference on Human Factors in Computing Systems 2005. pp. 471-480*

*Wagstaff, Jeremy (2007). The Anti-iPod: The Buddha Machine Shows That Bells and Whistles Aren't Always Better. Retrieved 9 November 2010 from Creative Deconstruction: http://online.wsj.com/article/SB117712103492177617.html*

*Walker, Rob (2007). Boxed Set. Retrieved 9 November 2010 from The New York Times: http://www.nytimes.com/2007/07/29/magazine/29wwln-consumed-t.html?_r=1*

## About the author

### Marc Hassenzahl

Has also published under the name of:
"Mark Hassenzahl", "Mare Hassenzahl", and "M. Hassenzahl"

Personal Homepage:
http://www.marc-hassenzahl.de

Marc Hassenzahl is Professor at the Folkwang University in Essen and research manager at MediaCity, Åbo Akademi University, Vaasa, Finland. He is interested in the positive affective and motivational aspects of interactive technologies – in short: User Experience

## YOUR NOTES AND THOUGHTS ON CHAPTER 3

Record your notes and thoughts on this chapter. If you want to share these thoughts with others online, go to the bottom of the page at: http://www.interaction-design.org/encyclopedia/user_experience_and_experience_design.html

**NOTES:**

# Chapter 4

# Social Computing

*by Thomas Erickson.*

As humans we are fundamentally social creatures. For most people an ordinary day is filled with social interaction. We converse with our family and friends. We talk with our co-workers as we carry out our work. We engage in routine exchanges with familiar strangers at the bus stop and in the grocery store. This social interaction is not just talk: we make eye contact, nod our heads, wave our hands, and adjust our positions. Not only are we busy interacting, we are also remarkably sensitive to the behaviors of those around us. Our world is filled with social cues that provide grist for inferences, planning and action. We grow curious about a crowd that has gathered down the street. We decide not to stop at the store because the parking lot is jammed. We join in a standing ovation even though we didn't enjoy the performance that much. Social interactions like these contribute to the meaning, interest and richness of our daily life.

VIDEO 4.1: Social Computing video 1 - Introduction to Social Computing.

*Courtesy of Rikke Friis Dam and Mads Soegaard. Copyright: CC-Att-ND (Creative Commons Attribution-NoDerivs 3.0 Unported). View full screen or download (9 MB). View transcription/captions.*

VIDEO 4.2: Social Computing video 2 - Three Best Practices and Three Major Pitfalls.

*Courtesy of Rikke Friis Dam and Mads Soegaard. Copyright: CC-Att-ND (Creative Commons Attribution-NoDerivs 3.0 Unported). View full screen or download (40 MB)*

Video 4.3: Social Computing video 3 - Face-to-face Interaction as Inspiration for Designing Social Computing Systems.

*Courtesy of Rikke Friis Dam and Mads Soegaard. Copyright: CC-Att-ND (Creative Commons Attribution-NoDerivs 3.0 Unported). View full screen or download (0)*

Video 4.4: Social Computing video 4 - Urban Planning as Inspiration for Designing Social Computing Systems.

*Courtesy of Rikke Friis Dam and Mads Soegaard. Copyright: CC-Att-ND (Creative Commons Attribution-NoDerivs 3.0 Unported). View full screen or download (0)*

Video 4.5: Social Computing video 5 - How to do Research in Social Computing.



Video 4.6: Social Computing video 6 - Social Computing: Visibility versus Privacy Manipulation versus Persuasion.

Video 4.7: Social Computing video 7 - How to integrate Social Computing in the Enterprise.

*Courtesy of Rikke Friis Dam and Mads Soegaard. Copyright: CC-Att-ND (Creative Commons Attribution-NoDerivs 3.0 Unported). View full screen or download (0)*

Video 4.8: Social Computing video 8 - The Reason IBM values Social Computing.

*Courtesy of Rikke Friis Dam and Mads Soegaard. Copyright: CC-Att-ND (Creative Commons Attribution-NoDerivs 3.0 Unported). View full screen or download (0)*

Video 4.9: Social Computing video 9 - Ethics in Social Computing.

*Courtesy of Rikke Friis Dam and Mads Soegaard. Copyright: CC-Att-ND (Creative Commons Attribution-NoDerivs 3.0 Unported). View full screen or download (0)*

## 4.1 SOCIAL COMPUTING: WHAT IS IT AND WHERE DID IT COME FROM?

Social computing has to do with digital systems that support *online* social interaction. Some online interactions are obviously social – exchanging email with a family member, sharing photos with friends, instant messaging with coworkers. These interactions are prototypically social because they are about communicating with people we know. But other sorts of online activity also count as social – creating a web page, bidding for something on eBay™, following someone on Twitter™, making an edit to Wikipedia[1]. These actions may not involve people we know, and may not lead to interactions, but nevertheless they are social because we do them with other people in mind: the belief that we have an audience – even if it is composed of strangers we will never meet – shapes what we do, how we do it, and why we do it.

Thus when we speak of social computing we are concerned with how digital systems go about supporting the social interaction that is fundamental to how we live, work and play. They do this by providing communication mechanisms through which we can interact by talking and sharing information with one another, and by capturing, processing and displaying traces of our online actions and interactions that then serve as grist for further interaction. This article will elaborate on this definition, but first let's look at where social computing came from.

The roots of social computing date back to the 1960's, with the recognition that computers could be used for communications and not just computation. As far back as 1961 Simon Ramo spoke of millions of minds connected together and envisioned "a degree of citizen participation ... unthinkable today." (Ramo 1961) Perhaps the best known vision is Licklider and Taylor's "The Computer as a Communications Device," in which they wrote of the development of "interactive *communities* of geographically separated people" (Kittur and Kraut 2008) organized around common interests and engaging in rich computer-mediated communication.

The first general purpose computer-mediated communication systems emerged in the 1970's. Examples include Murray Turoff's pioneering EMISSARY and EIES systems (Hiltz and Turoff 1993) for "computer conferencing," PLATO Notes at the University of Illinois (Wooley 1994), and the first mailing lists on the ARPANET. Others followed and the 1980's saw a flowering of online systems that supported social interaction via online text-based conversation: bulletin board systems, Internet Relay Chat, USENET, and MUDs (see Howard Rheingold's *The Virtual Community* for a good history (Rheingold 1993)). The early 1990's saw continued improvements in basic communications technology – speed, bandwidth and connectivity – and the advent of the Web. Although initially the Web only weakly supported social interaction by allowing people to display content and link to web pages of others (Erickson 1996), it marked the beginning of the widespread authoring and sharing of digital content by the general public.

In my view, social computing came into its own in the late 1990's and early 2000's when digital systems became capable of doing more than simply serving as platforms for sharing online content and conversation. The key development was the ability of digital systems to process the content generated by social interaction and feed the results of that processing back into the system. That is, while computer conferencing and its successors served as platforms that supported the production of vast tracts of online conversation, the conversation itself was understood *only* by humans. Digital systems provided passive media through which people interacted. The advent of modern social computing came when digital systems began to process user-generated content and make use of it for their own purposes – which often involved producing new functionality and value for their users.

A good example of creating value by processing user-generated content is Pagerank™, the algorithm used by the Google[1] search engine. The fundamental insight of Pagerank is that the importance of a web page can be estimated by looking at the number of pages that point to it (weighted by the importance of those pages, which can be recursively evaluated in the same way). The underlying assumption is that the act of creating a link to a page is, on the part of a human, an indication that the page is important in one way or another. Thus Pagerank mines and aggregates the results of human judgments as expressed through link creation, and uses it to assess the importance of pages and determine the order in which to display them. This is an early, and very notable, example of the recognition that the digital actions of a large number of people can be tapped to provide a valuable service.

Let us pause, and summarize what we've covered so far. Social activity is a fundamental aspect of human life. Not surprisingly, digital systems have accommodated such activity for decades, initially serving as platforms that supported online conversation and other collaborative activity. More recently an important shift has occurred: systems have become able to make use of the products of their users' social activity to provide further value, and that in turn amplifies the use of the system.

## 4.2 AN EXAMPLE: A SOCIAL COMPUTING MECHANISM

This is abstract, so let us look at a common example, that of the online retailer Amazon.com[1]. As most readers will be aware, Amazon is an online department store that sells a wide variety of goods, as well as providing an online storefront for other retailers. While there is nothing in Amazon's core business – selling goods online – that requires social computing, Amazon is notable because it used social computing mechanisms to differentiate itself from other online stores. For illustrative purposes, we will take a close look at Amazon's product review mechanism.

Amazon enables its users to create online product reviews. Each review consists of a textual essay, a rating of 1 to 5 stars, and the name of its author. Products may garner many reviews – for example, the best-selling book, *The Girl with the Dragon Tattoo*, has amassed over 2500 reviews. If this were the extent of Amazon's review mechanism, it would be interesting and useful, but not a social computing mechanism: it would be akin to the early systems that served as platforms for producing user-generated content that was only understood by users. Like those systems, while it is valuable to provide a large number of user-generated reviews, it seems unlikely that viewers will really read through all 2500+ reviews of *The Girl with the Dragon Tattoo*.

What makes the difference, and moves Amazon's review mechanism into social computing territory, is that Amazon has been clever about the kind of information it enables its users to enter. Besides the 1-5 star ratings of books from which it can produce averages and other statistics, it allows users to essentially review the reviews: readers can vote on whether a review is "helpful" or not, can flag it as "inappropriate," and can enter a comment. And readers, in fact, do this. The "most helpful" review of *The Girl with the Dragon Tattoo* has been voted on by over 2,000 readers, and has received 44 comments.

This is significant because users' votes and ratings can be used (just as Pagerank uses links) to provide valuable information. Using this information, Amazon provides two user interface components that significantly increase the utility of the user-entered information (Figure 4.1). The first is a graph of the distribution of a book's ratings that provides viewers with an at-a-glance summary of reviewer sentiment. The second is that it uses the number of "helpful" votes to foreground particular reviews – e.g., the most helpful critical review and the most helpful favorable review.

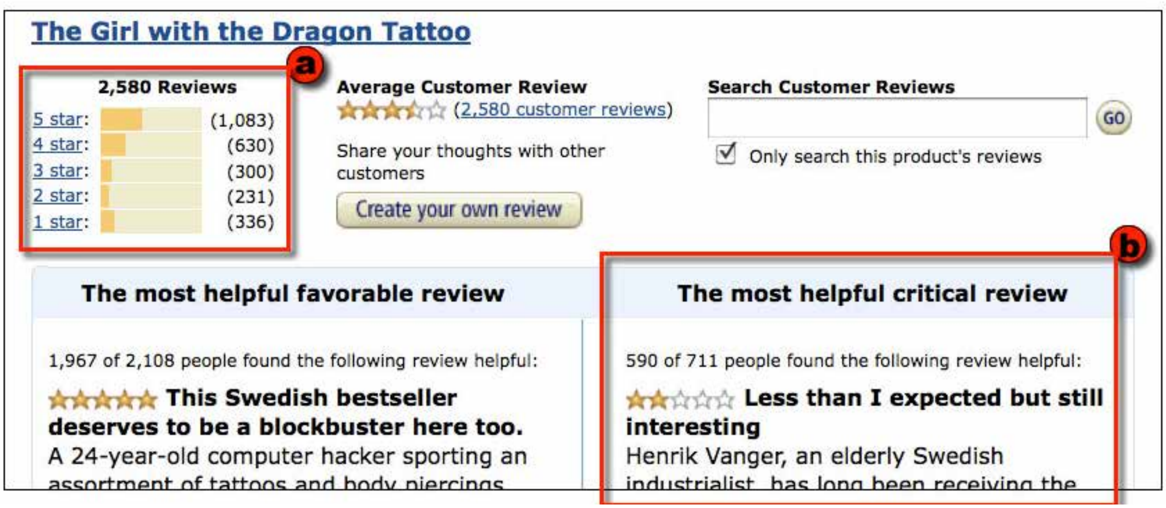



FIGURE 4.1: The Amazon.com book review summary includes components that show (a) the overall distribution of review ratings, and (b) the "most helpful" (in terms of readers' votes) critical review.

These components – which rely on computations carried out on user input – make it easier for viewers to deal with large amounts of user-generated content. The first integrates the results of all reviews of a book, providing not only the average rating but also the more informative distribution of ratings. The second uses the review's book ratings in tandem with helpful votes to highlight particular reviews – the most helpful favorable review, and the most helpful critical review. Now, rather than wading through dozens, hundreds or thousands of reviews, the viewer can glance at the overall distribution and read through the "most helpful critical review." This increases the useful of Amazon's review information, and most likely increases visits by prospective purchasers. In addition, the possibility of "helpful" votes, and the chance to be recognized as the author of a "most helpful" review, may serve to incent reviewers to write better reviews. All in all, these mechanisms produce virtuous circles: positive feedback loops that promote desirable results.

This aptly illustrates the phase shift that began around the year 2000. Systems emerged that were more than platforms for social interaction: the results of users' activities began to be usable not just by users, but by the digital systems that supported their activity. Sometimes ordinary content is made digitally tractable by dint of computation, as with Google's Pagerank algorithm that mines the web to determined linked-to frequency. (Amazon takes this approach as well, when it uses the purchase history of a user to identify those with similar histories, and then provides users-like-you-also-bought recommendations). And sometimes the system requests that users directly enter data in a form that it can make use of – like Amazon's ratings and "helpful" votes, or the "I like this" and "favorite" votes used in other systems. However it occurs, this ability for the information produced via social interaction to be processed and re-used by the system supporting that interaction is the hallmark of present day social computing.

## 4.3 THE VALUE OF SOCIAL COMPUTING

Why does social computing matter? Besides the fact that the social interaction supported by social computing systems is intrinsically rewarding, there are a number of ways in which social computing systems can provide value over and above that offered by purely digital systems.

First, social computing systems may be able to produce results more efficiently. Because Amazon can draw on its entire customer base for book reviews, it can provide far more reviews far more quickly than relying on the comparative trickle of reviews produced by *Publishers Weekly* and other trade outlets. *The Girl with the Dragon Tattoo* received five reviews within a month of its 2008 publication in English, long before it emerged from obscurity in the English language market. Similarly, Wikipedia, the online encyclopedia, offers over three and a half million articles in the English edition, and can generate articles on current events literally overnight. For example, within an hour of the 2011 Tōhoku earthquake and tsunami in Japan, a three paragraph article had appeared; that, in turn, was edited over 1,500 times in the next 24 hours to produce a well-formed article with maps, photos and 79 references. (As of this writing, nearly ten weeks after the event, the article has been edited over 5,100 times by over 1,200 users, and has 289 references; in the last 30 days it has received nearly 600,000 views.)

A second way in which social computing can be of value is by increasing the quality of results. A good example is the MatLab[1] open source programming contest (MatLab Central 2010, MatLab Central 2011). MatLab is a commercial software package for doing mathematical analysis that uses its own scripting language, and one way its developers promote it is by running a programming contest. Each contest poses a problem, and the challenge is to produce a program that solves it as quickly and completely as possible. Contestants submit the MatLab source code of their programs, and the programs are immediately evaluated, scored, and listed in order of their scores. What makes the contest unusual and interesting is that each entry's source code is also made publicly available. So when a new top-scoring program is submitted, other contestants will download it and look through the source code for ways to "tweak" it so that it runs slightly faster. The new tweaked program can be submitted, and it and its author will thereby vault ahead of the originator into first place (until another contestant tweaks *that* entry). As this tweaking process is repeated dozens of times (e.g., Figure 4.2), the entry is rapidly optimized, drawing on and integrating the specialized knowledge of the community of MatLab programmers (Gulley 2004).

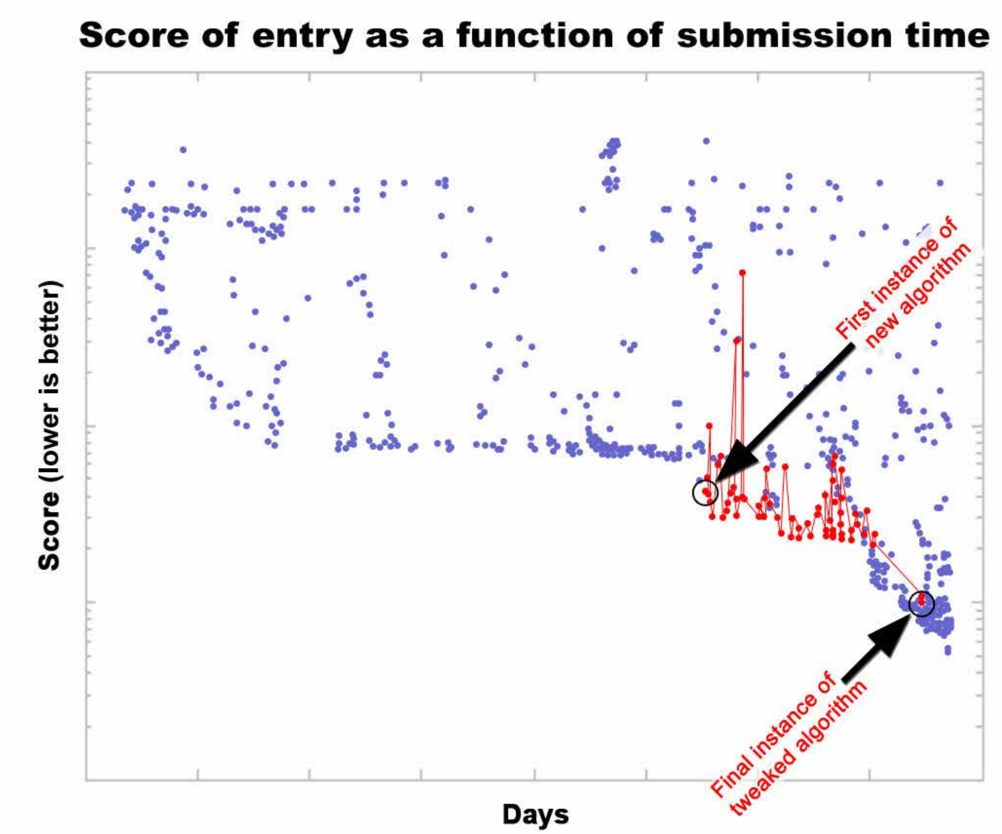



FIGURE 4.2: Collective optimization in the MatLab open source programming contest. Following the introduction of a new algorithm (variants shown in red), contestants refine it, gradually (and noisily) optimizing the algorithm (adapted from MatLab Central 2010).

A third way in which social computing systems can provide value is by producing results that are seen as fairer or more legitimate. Thus, to return to Amazon, one might trust the results of 'the crowd' of reviewers more than an 'official' reviewer who may have values or biases that are out of sync with the ordinary reader. Another example is the online auction, where multiple people bid for an item – those who lose out may not like the result, but few will argue that it is not legitimate. Stepping out of the digital realm for a moment, note that the rationale for the plebiscites on which democracies are based is not that they produce more rapid decisions, nor that the decisions are necessarily of higher quality, but rather that they are representative and reflect the popular consensus. It is notable that the value of plebiscites and auctions (and even the Amazon review process) can be invalidated by failures in their processes – ballot box stuffing, vote buying and other forms of fraud in elections; shills in auctions; and collusion among bidders and reviewers. In cases like this, it is the *legitimacy* of the result that has been undermined; demonstrating that the decision was arrived at more quickly or is of higher quality is immaterial. In this case the value of the product is contingent upon the process through which it was derived.

A fourth way in which social computing provides value is by tapping into abilities that are uniquely human. For example, the ESP Game (Ahn and Dabbish 2004), which we will discuss in more detail shortly, is an online game in which a user and an anonymous partner look at an image and try to guess the words that occur to the other person. Both enter words simultaneously, and when they both enter the same word they 'win' and receive points; a side effect of this is that the players are producing textual labels for the image – a task that, in general, computer programs cannot perform. Other examples are Galaxy Zoo (Galaxy Zoo 2011, Priedhorsky et al 2007), which asks people to classify galaxies in astronomical photographs by their shapes, and Investigate Your MP (Guardian - guardian.co.uk 2011), which asks participants to read through politicians' expense reports and flag those that seem suspicious.

To sum up, there are different ways in which social computing systems may produce value: they may produce results more quickly by multiplying effort; they may produce higher quality results by integrating knowledge from multiple participants; they may produce results that are more legitimate by virtue of representing a community; and they may carry out tasks that are beyond the capacity of current digital systems by drawing on uniquely human abilities. But while this value is of great practical import, it should not obscure the most important aspect of social computing: the social interaction itself. Greater efficiency, quality and legitimacy are important benefits, but the reason most people engage with social computing systems lies in the give and take of the interaction itself, the meaning and insight we derive from it, and the connections with others that may be created and strengthened as its result.

## 4.4 SOCIAL COMPUTING AS A SYSTEM: THE ESP GAME

Thus far we've introduced the notion of social computing as an approach that does more than provide a platform for social interaction – it makes use of social interaction to produce various forms of value. The shift to social computing is, at the heart, driven by the ability of digital systems to process the products of the social interaction they support. The products of social interaction have been made digitally tractable, either by dint of digital computation (e.g., Pagerank), or by persuading users to enter information in a form that the digital system can use (e.g., Amazon's "Helpful" votes and Five-star ratings).

Up to this point our principal example of social computing has been Amazon. However, while Amazon has been enormously successful at making use of social computing mechanisms, if one removed all elements of social computing from Amazon, it would still be able to carry out its basic aim of selling goods online. To expand our understanding of social computing we'll take a look at some examples of social computing *systems* – that is, systems that, without social computing mechanisms, simply would not function at all.

The ESP Game (Ahn and Dabbish 2004) is one of a class of systems that have been characterized as performing "human computation." This type of system is designed so that it enables a large number of people to perform a simple task many times (and often many many times). The art of designing this type of social computing system lies in finding a domain with a difficult problem that can be solved by the massive repetition of a simple (for humans) task, and in figuring out how to motivate the human participants to carry out a simple task many times.

The ESP Game is notable both for its practical success and for the subtleties of design that underlie its apparent simplicity. At a high level the ESP Game sets out accomplish the task of assigning textual labels to images on the web. This is a task that is difficult for computers to perform, but easy for humans. However, while easy for humans, it is not a task that is very *interesting* to perform, which given that there are billions of images in existence constitutes a problem. What the ESP Game does is to reframe the image labeling process as a game, and by making it fun it succeeds in recruiting large numbers of people to label images. In fact, in its first 5 years of existence, 200,000 people used it to produce more than 50 million image labels (Ahn and Dabbish 2008).

The ESP Game works as follows. A user goes to the web site, where, after a brief wait, he or she is paired with an anonymous partner and the game begins (Figure 4.3). Both participants are shown (a) an image, and are asked to type (b) words that they believe their partner will type; they may also be shown (c) "taboo words" which cannot be used as guesses. When they achieve (d) a match, they receive points, and move on to a new image; if it proves too difficult to achieve a match, a player can click (e) the "pass" button, which will generate a new image. Each game lasts three minutes, and both participants receive points each time they match on a word.

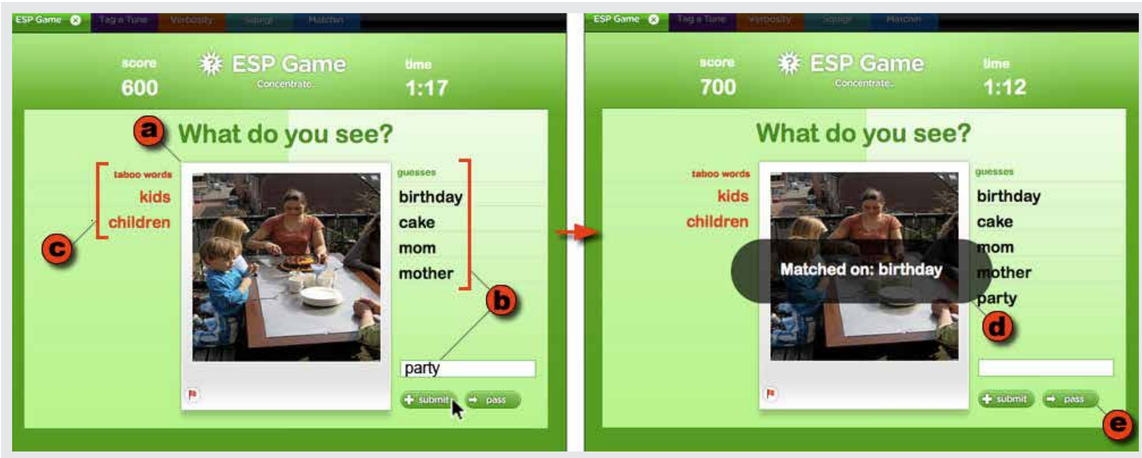



Figure 4.3: The ESP Game in Play mode. The player (a) looks at an image, (b) enters words that describe the image, except that (c) certain words called "taboo words" can't be chosen. If the other player enters one of the same words, there is (d) a match, and both players get points. If it seems too difficult to come up with good guesses, either player can (e) pass.



After a game ends the players go their separate ways, each seeing a screen (Figure 4.4) that recaps how well they did, both in the individual game, and in a cumulative score for all games played. Players are also shown how far they have to go to achieve the next "level," and how they compare with the highest scoring player for the day.


FIGURE 4.4: The ESP Game Score window. At the end of the game, players are shown (a) their points, (b) their level, (c) points needed to achieve the next level and (d) to beat the best player of the day (d). Players can also earn points by (e) referring friends.



The ESP Game has a number of design features that illustrate issues that social computing systems, in general, must address. As we shall see later, different systems may address these issues differently, but the ESP Game provides a good starting point for grounding this discussion.

### 4.4.1 Computation

Social computing systems carry out various forms of work to produce value, often by applying algorithms to the results of user-generated content. The ESP Game performs computations by incenting individuals to use their perceptual and cognitive abilities to generate possible labels for an image, and aggregating results across many games to produce a valuable outcome. The result of a single game is a set of images each of which has either a word that both players typed in response to the image, or a "pass." As multiple pairs play the game on the same images, sets of labels are produced for each image. The best labels will be those that are produced most frequently, and after a while the game will add them to the list of "taboo words" that are not allowed as guesses. This requires players to produce less obvious labels, which are in turn added to the "taboo word" list, until it becomes sufficiently difficult that when presented with an image and its list of taboo words, most players pass. At this point, the image can be 'retired' from the game and considered to have a complete set of labels weighted by the frequency with which people produced them. This is a result that cannot be achieved by purely digital systems.

### 4.4.2 Recruiting and Motivating

The ability of a social computing system to produce value relies on user-generated content, and that means that the system must take measures to ensure that it has a sufficient number of users who are motivated to participate. This was not an issue for Amazon, because the Amazon review mechanism is embedded in the larger Amazon ecosystem, and it happens that some of those attracted to Amazon by its function as an online retailer are interested in reviewing products. This is not the case with the ESP Game – it must do all the work of attracting people. It does this via its use of game-like incentive mechanisms to recruit and motivate its players. People hear about the ESP Game via word of mouth – players can earn points by referring others – and come to play it because it's fun. Once potential players arrive at the site the problem shifts to engaging them in the game. To that end the ESP Game is nicely designed with bright colors, snappy interaction and appropriate sounds. Many of its features – the limited time, the awarding of points for right answers, a graphical scale showing cumulative points, and the sound of a clock ticking during the final moments of the game – work to motivate the users during game play. When a game ends, other features – cumulative points, user levels, points needed to achieve the next level, and points needed to beat the top player of the day (Figure 4.4) – encourage the player to "play again!!!" This also provides motivation for players to register, creating online identities that can accumulate points across sessions, and that can vie for positions in the top scores lists for "today," "this month," and "all time." All of these features serve to engage users, motivate them, and encourage them to return – issues that any social computing system will need to address.

### 4.4.3 Identity and Sociability

Not only must social computing systems attract and motivate their users, but they must make them 'present' within the system. Participants in a social computing system generally require identities through which to engage in interaction with others, and identity – especially identity that persists over time – is also bound up with motivation and reputation. The ESP game is actually a relatively low-identity example of a social computing system, in that its participants are not allowed to talk with one another while playing the game, so as to deter cheating. Nevertheless, the ESP Game does take pains to support identity and reinforce the social aspects of the game. As noted, players can register, creating a screen name, an icon, and other elements of a profile. While communication between a pair of players during the game is prohibited, players can join a chat room for the site as a whole (the ESP Game is part of a site called *Games with a Purpose*). More generally, the design shows the presence of others. Once a user chooses to play, there is a brief wait while the game is "matching you with a partner." Once a match is made, the player is told the screen name and shown the icon of their partner. Like the incentive mechanism, these social features aim to increase the attraction and interest of the site.

But suppose you show up for a game and there is no one to play with? This is a problem in that not only can the game not take place, but the player who has come to the site may now be less likely to return. The ESP Game deals with this situation by using autonomous software programs called "bots." If a visitor arrives at the ESP Game site and no one else is there, the visitor will still be paired up with another "player," but unbeknownst to the visitor the other player will be a bot. The game that ensues will use images that have already been labeled by at least one pair of human players, and the bot will simply replay the responses (and timings) of one of the previous players, giving the human partner the experience of playing against someone else. This use of bots supports the experience of the game, and has another use that we will look at shortly.

### 4.4.4 Directing and Focusing Activity

Another issue that social computing systems need to deal with is how to focus or otherwise shape the activities of their users. In the ESP Game, this is done via taboo words. As already described, taboo words serve to increase the breadth of the set of labels generated for an image by ruling out those that many previous pairs of players have produced. Taboo words also shape the set of labels produced in a more subtle way: they can prime players to pay attention to certain aspects of the image (Ahn and Dabbish 2008) (e.g., an image with "green" as a taboo word might incline players to name other colors in the image). The ESP Game could use other approaches to focusing work such as selecting images from particular known sets (e.g., images of paintings), or recruiting players from particular populations (e.g., art school students). Many social computing systems have mechanisms, of one sort or another, that try to focus or otherwise control the nature of the computation the system performs.

### 4.4.5 Monitoring and Controlling Quality

While humans can perform computations that are difficult or impossible for digital systems, it is also the case that human-generated results may be inaccurate – thus many social computing systems need to address the issues of monitoring and controlling the quality of results produced. Quality problems may result from ignorance, unnoticed bias, or intentional choice. In the case of the ESP Game, the primary threat to quality is cheating. That is, the game-like incentive mechanisms work so well that players may play with the goal of getting points, rather than accurately labeling images.

As the ESP Game has developed, various cheating strategies have been identified and circumvented. Solo cheating occurs when a person logs on twice and tries to play themselves – this can be detected and prevented by IP matching. Dyadic cheating occurs when two players devise a word entry strategy (e.g., “one”, “two”, “three”) and try to log on at the same moment in the hopes of being paired up – this can be prevented by having a long enough waiting period (“matching you with a partner”) and a sufficient number of waiting players that it is unlikely that conspirators will be matched. If there are not enough players waiting to ensure a good likelihood of a random match, the ESP Game can use bots as surrogate players, as previously described. Finally, cheating can occur *en mass* when someone posts a word entry strategy and starting times on a public web site. This approach can be detected by a sudden spike in activity (or a sudden increase in performance), and countered by, once again, pairing players with bots.

These examples of cheating raise several points. First, with respect to designing social computing systems, cheating can be dealt with. It is simply necessary to identify cheating strategies and block them – or at least lower their probability of success to a point where it is easier to win by using the system as the designers intended. Second, note that cheating is an issue only in certain types of social computing systems. Cheating occurs primarily in systems where the incentive mechanism is unrelated to the system’s purpose. Third, note that since cheating removes the fun from the game, its existence is apt testimony to the power of the ESP Game’s game-like incentive mechanisms.

### 4.4.6 Summary

In this section we’ve looked at social computing systems as systems, using the ESP Game as an illustrative example. Unlike the Amazon review mechanism, which was embedded in the larger Amazon ecology, the ESP Game needs to function as a complete system, solving the problems of recruiting participants, giving them an identity within the system, focusing their attention on tasks that need doing, incenting them to do the task, and monitoring and controlling the quality of the results. The ESP Game does this by drawing on game design thinking. It is successful because the tasks on which it is focused are simple, well-formed and thus amenable to very rapid, very iterative interaction – and this, in turn, is well suited to game play.

On the other hand, while von Ahn and his colleagues have proven to be quite ingenious in their ability to find domains amenable to this approach (see Ahn and Dabbish 2008), many problems do not break down so neatly into such simple well-formed tasks. Yet, as we shall see, social computing systems – albeit with different approaches to the above issues – can still make headway.

## 4.5 SOCIAL COMPUTING AS A SYSTEM: WIKIPEDIA

In this section we examine what is, in the view of many, the most successful example of a social computing system: Wikipedia. Besides its success Wikipedia is of interest because it offers a stark contrast with the ESP Game. Whereas the ESP Game attracts a steady stream of anonymous users who perform a simple task embedded in a game, Wikipedia is more of a community, with a core of committed participants who interact with one another while performing a variety of complex tasks. Wikipedia has also proved to be popular with researchers, making it a superbly studied example of a social computing system. Thus, our examination of Wikipedia will add breadth to our understanding of social computing.

As most readers will know, Wikipedia is, in the words of its slogan, “the free encyclopedia that anyone can edit.” With a few exceptions, every article in Wikipedia has an “edit” tab that allows anyone to edit the article (Figure

4.5, b). On the face of it, this is a bit of a paradox: how can one have an authoritative source of knowledge that anyone can change at any moment? And yet it works well enough. While generalizing about millions of articles in all stages of development is difficult, it is fair to say that Wikipedia's accuracy is surprising. Studies have shown that some classes of articles are comparable in accuracy to their counterparts in the Encyclopedia Britannica (Giles 2005), and, more generally, that an article's quality tends to increase with the number of edits it has received (Wilkinson and Huberman 2007, Kittur and Kraut 2008).

Regardless of how it compares to the quality of traditional encyclopedias, Wikipedia has been remarkably successful. With over three and a half million articles in the English edition alone, it is among the most visited sites on the web. And, as we saw earlier, it can generate lengthy, well-researched articles very quickly – literally over night at times. As encyclopedias go, this puts Wikipedia in a class by itself.

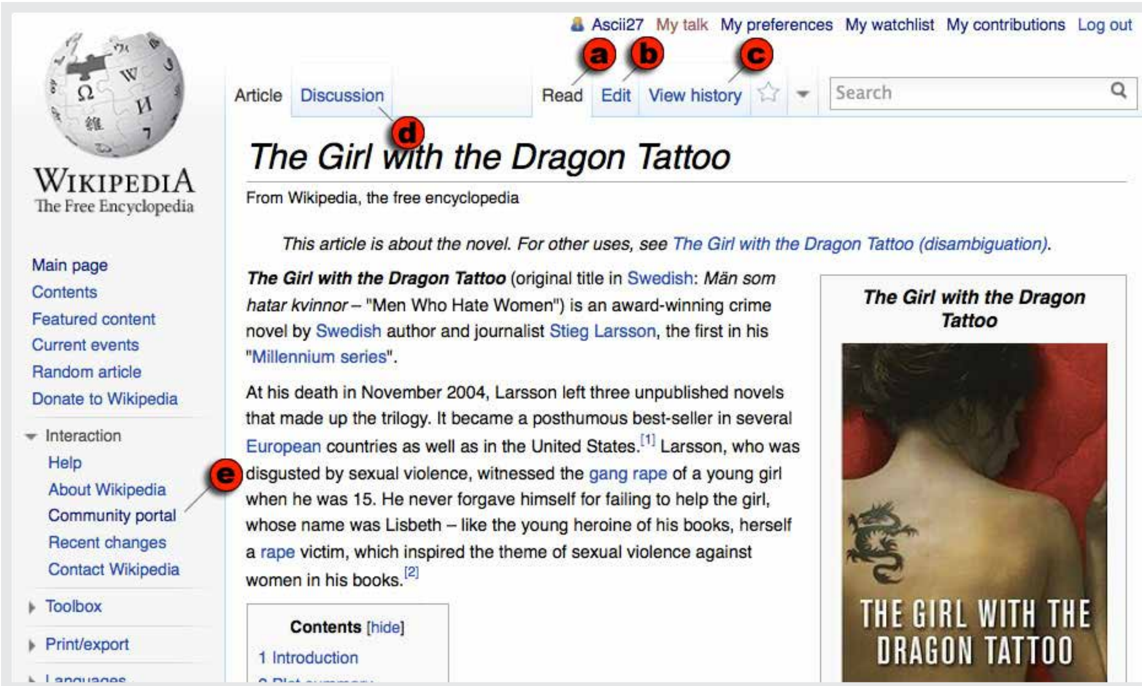



FIGURE 4.5: The Wikipedia article page for *The Girl with the Dragon Tattoo*. Most visitors come (a) to read, but they can also (b) edit the article, (c) view its history, or (d) read its discussion page. Those wishing more involvement in Wikipedia can visit (e) the Community portal.



## 4.5.1 Contributing to Wikipedia

Let's begin our examination by considering what contributors do in Wikipedia. While the aim is to create an encyclopedia article, clearly this is too large a task. Contributors do not author fully formed articles all at once. Instead, articles coalesce out of the accretion of smaller efforts. One contributor writes an article "stub," others add paragraphs, and still others expand, modify and condense existing text. Some may add links to references, others may correct typos and grammatical errors, and still others may contribute images. This is how the article on *The Girl with the Dragon Tattoo* developed. Starting from an article stub that appeared at about the time of the book's English language publication, the article grew gradually – with bursts of activity when movies were released – until today it is a well-formed article that has been edited over 600 times by 395 contributors. As of this writing, it has been viewed 234,000 times in the last 30 days.

The key question to be asked about Wikipedia is this: How it is that Wikipedia articles improve in quality over time? How is it that Wikipedia determines that a particular change – whether it's replacing one word, adding a paragraph, or reorganizing an article – is a change for the better? Sometimes it's obvious – for example, correcting a typo – but more often than not it isn't obvious. The answer is that Wikipedia relies on users to judge the quality of changes to an article. But this is not much of answer. What is important is – and what constitutes the art of how Wikipedia is designed – is the way in which it supports its users in making such judgments. As we shall see, Wikipedia musters a complex array of social computing mechanisms in support of the activities of its contributors.

### 4.5.2 Judging Quality and Making Changes

Two things have to happen for users to be able to judge the quality of a change: individual changes must be made visible; and users must be able to express their opinions on the desirability of a change. Wikipedia accomplishes this through its revision history mechanism (Figure 4.6) that is accessed by the "View History" tab on each article. The revision history lists all the changes made to an article and provides a link that enables the viewer to undo the change. Thus, if the entire text of the article has been replaced with a string of obscenities – an action more frequent than one might expect, and referred to as "vandalism" – the viewer can click on an "undo" link and revert the article to its prior state. And indeed, one of the early and surprising research findings about Wikipedia was that such acts of vandalism were typically discovered and reverted within two to three minutes (Viégas et al 2004) this result has continued to hold up over time (Priedhorsky et al). This becomes less surprising in view of the fact that Wikipedia provides a mechanism called a "watchlist" that allows users to monitor changes to articles they care about. For instance, the article for *The Girl with the Dragon Tattoo* currently has 59 people watching it.

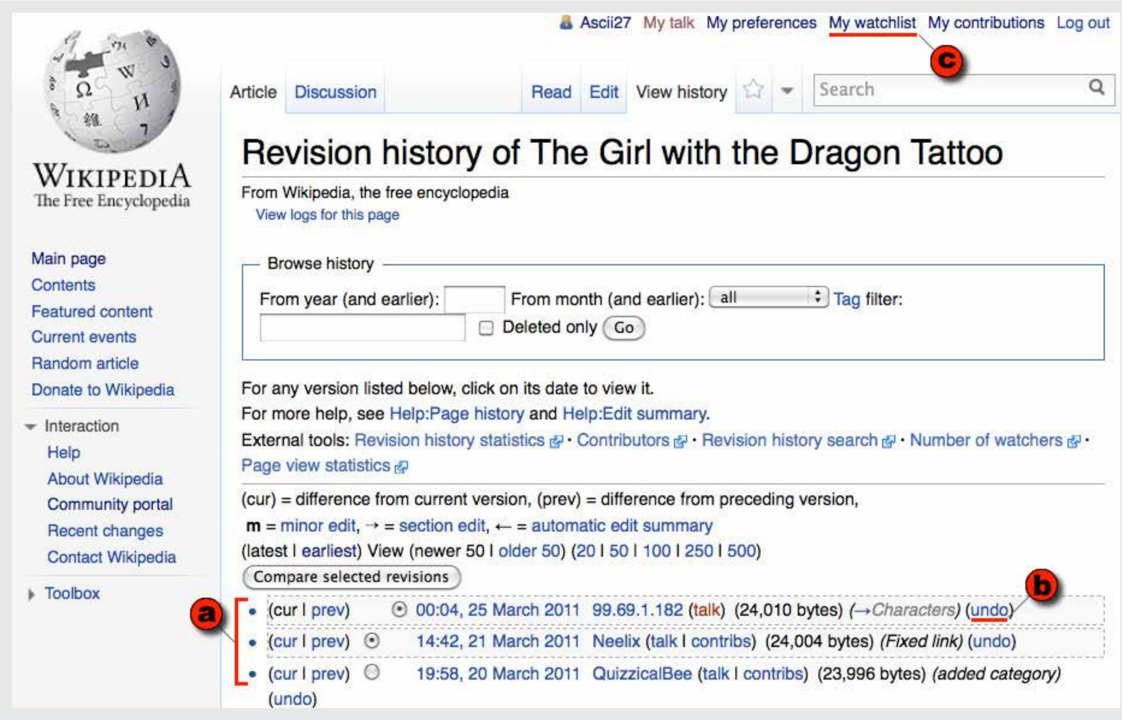



Figure 4.6: A Wikipedia revision history page. The revision history shows (a) a list of all changes made to an article, (b) provides a way to undo each change, and (c) enables those who care about an article to add it to a "watchlist" so changes to it can be watched.



However, there is more to supporting quality than detecting and undoing vandalism. After all, vandalism is obviously a change for the worse. Much vandalism is trivial to detect. The more difficult issue is how to resolve subtler questions such as whether an explanation is clear or obscure. Or whether the reorganization of a paragraph improves it. Or whether a particular picture is helpful. Or whether a way of describing something departs from the neutrality desirable in an encyclopedia. To make decisions about these types of quality issues people need to communicate – and the revision history page lays the foundation for such communication. To see this, take a look at the section of the revision history shown in


Figure 4.7: A segment of a Wikipedia revision history showing four revisions. For each revision there is (a) a way of comparing to other revisions, (b) a time and date stamp, (c) links to the user who made the change, (d) information about the change, and (e) a way to undo the change.



The revision page contains a list of every change made to its article. Each entry in the list contains (a) a way to compare that change to other versions of the article, (b) the time and date of the change, (d) other information about the change), and (e) a way of undoing the change. Of particular interest for our purposes is that the entry also contains (c) information about *who* made the change. Specifically, the entry contains

- the name or IP address of the person who made the revision
- a link to that person's talk page (which provides a way to communicate with them), and
- a link to a list of contributions that person has made to Wikipedia (which provides a way to judge how experienced they are)

This expands the revision history into a social mechanism by introducing identity and communication channels. That is, rather than just seeing a change and deciding whether to revert it, the viewer can see who made the change, find out something about the user (via the link to their user page), their experience with Wikipedia (via the link to their contributions page), and even discuss their change with them (via the link to their talk page). Indeed, researchers have shown that the quality of Wikipedia articles is not simply related to the number of people who edit them, but that for this relationship to hold these contributors must also be engaging in communication and collaboration (Kittur and Kraut 2008, Wilkinson and Huberman 2007). For example, investigating the first entry in Figure 4.7, one can quickly discover that Varlaam is a highly experienced Wikipedia editor who has been awarded a "Master Editor" barnstar and has contributed to dozens (at least) of articles on books and films. Even if one disagreed with Varlaam's change, one might hesitate to simply undo it, given the level of experience in evidence. The second entry is a different story – it has only an IP address associated with it, which means that this person has not registered with Wikipedia. Nevertheless, by clicking on the IP Address link, one can quickly see that the person at this IP address has been making regular contributions to Wikipedia over the last six months on a variety of articles related to films, and one can even take a look at individual contributions and see that this person has been a positive contributor.

### 4.5.3 Identity and Communication

The links in the revision history page illustrate another aspect of Wikipedia: it has a variety of mechanisms that support identity and communication. Every person who contributes to Wikipedia has a "user page," a "contributions page," and a "talk page." The user page is like a home page, where a contributor can post whatever they like about themselves. Often this will include information that tells others about their experience and knowledge *vis a vis* topics they like to edit. This page is also where Wikipedia contributors display awards they've received from the Wikipedia community (Wikipedia has a custom of encouraging contributors to give symbolic awards – the best known example being a "barnstar" – to others who have helped on an article or project). In addition to the user page, there are two other automatically generated pages. The contribution page lists every change that that person has made to Wikipedia, and includes a "diff" link that shows precisely what changes the user has made. And the talk page supports conversation with that user. Both the talk and contribution pages have direct links from the entries in the revision history. In addition to *user* "talk pages," there are also *article* "discussion pages" for discussions about the content of an article. For example, the discussion page for *The Girl with the Dragon Tattoo* has seven discussions, one on whether it is legitimate to characterize the protagonist as "incorruptible," another about why the English title is so different from the original Swedish title, and yet another about whether to include a claim about an authorship dispute.

While all of these ways of finding out what people have done and talking with them might seem like a recipe for chaos, overall it appears to work. People *do* talk with one another, and their discussions tend to be about whether and how to change the content of the articles. For instance, more than half of the comments on talk pages are requests for coordination (Viégas et al 2007). And although argument is common, contributors often reach a rough consensus. What is interesting about this is not that people reach consensus, but rather how they reach agreement. Wikipedia has an extensive set of policies and guidelines that govern it (Figure 4.8a). For instance, one of Wikipedia's fundamental principles is that articles should strive for a neutral point of view, and try to fairly represent all significant views that have been published by reliable sources. Another policy is verifiability – that is, that readers should be able to check that material in Wikipedia that is likely to be challenged must be attributed to a reliable source through an in-line citation.

What is important here, for the purposes of understanding social computing, is *not* the policies and guidelines themselves, but rather that policies and guidelines function to provide an infrastructure for discussion. Ideally, contributors who differ argue about whether something is in accord with the policies or not, rather than attacking one another. Or contributors discuss and evolve the policy itself. Just like Wikipedia articles, Wikipedia policies have their own pages with "Edit," "Discussion," and "View History" tabs, and, just like articles, policies and guidelines are extensively discussed and developed. Thus the article on the *Neutral Point of View* policy (Figure 4.8b) has seven sections that cover about 7 pages (4,300 words), and has been edited by over 1700 users over the decade it has been in existence; it has been viewed over 37,000 times in the last 30 days which, while not nearly as much as *The Girl with the Dragon Tattoo*, nevertheless indicates that it is an actively used resource. Both the use of policies to guide editing, and the collective evolution of policies by the community of users are discussed at length in the aptly named article, *Don't Look Now, But We've Created a Bureaucracy* (Butler et al 2008).

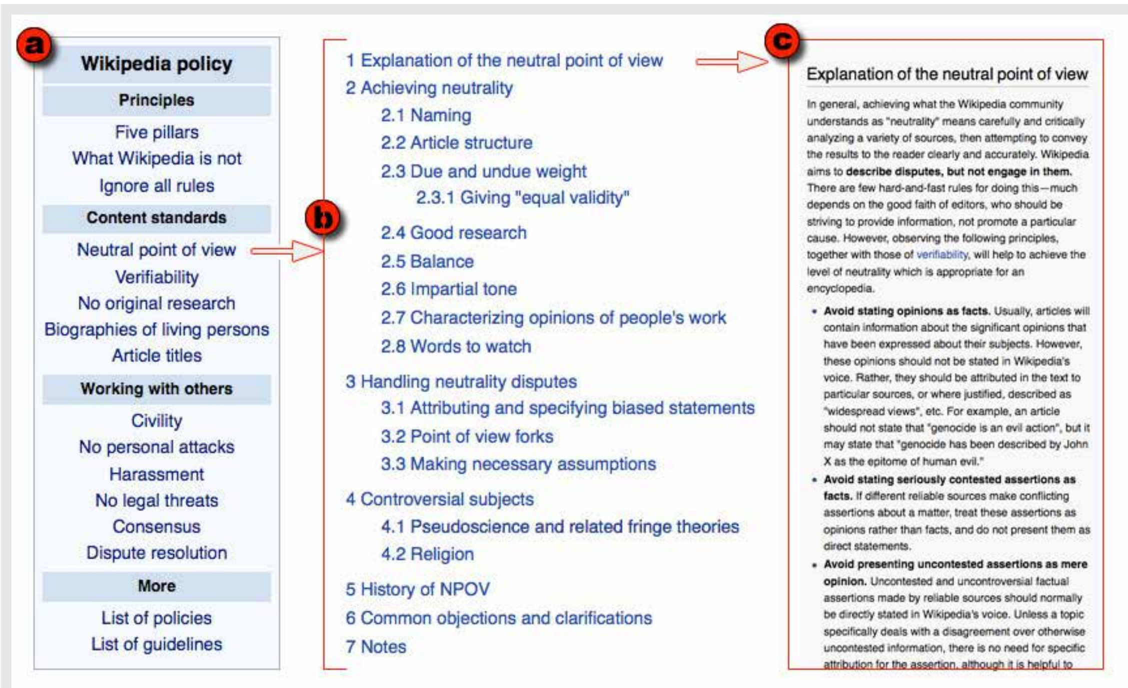



FIGURE 4.8: Wikipedia has a well-articulated set of guidelines and (a) policies, each of which is embodied as a Wikipedia article with (b) a detailed outline and (c) a thorough discussion, and which can be revised in the same way as ordinary articles.



So what we've seen here is that although random people can and do click the edit tab and make changes on the spot, much of the work that happens in Wikipedia does not play out so simply. Mechanisms that support identity, communication, and the application of policies come together to enable a complex social process the guides the creation and development of high quality articles.

Given this, it is perhaps not surprising that there is a core of contributors, often referred to as Wikipedians, who are responsible for the majority of contributions. While there are various ways of defining what counts as a contribution, researchers agree that a small percentage of editors contribute the majority of content – for instance, the top 10% of contributors by number of edits provide between 80 and 90% of the viewed content (Priedhorsky et al 2007). More generally, Wikipedians do more work, make contributions that last longer (before someone else changes them), and invoke community norms to justify their edits more frequently (Panciera et al 2009). In sum, although Wikipedia "is the free encyclopedia that anyone can edit," not just anyone tends to edit it, and there is much more to "editing" than meets the eye.

### 4.5.4 Summary

Wikipedia is a remarkable achievement. It is a self-governing system that produces well-structured articles – sometimes literally over night – that are sufficiently useful that it is among the most visited sites on the web. In our examination of Wikipedia we asked how it is that Wikipedia articles improve over time, noting that, for example, obvious problems like vandalism are repaired within two to three minutes. In what should by now be a recognizable motif, we saw that Wikipedia, as a system, 'knows' something about its content. Specifically, Wikipedia keeps track of every change ever made to every article, and it makes those individual changes visible and actionable on the revision history page of each article. Wikipedia (or, strictly speaking, the design of Wikipedia) recognizes that some changes are worth keeping, and others are not, and that by making it easy for users to view and pass judgment on those changes it can support the creation of increasingly higher quality articles. And because quality can be a subtle and contentious issue, Wikipedia provides ways for users to talk with one another about changes, and provides policies that guide users in making consistent decisions about those changes. In this we see the *modus operandi* of social computing: users add content, and the system processes that content in ways that make it more usable (in this case, by increasing the ability of people to discuss, evaluate and keep or undo changes in keeping with Wikipedia policy).

## 4.6 SOCIAL COMPUTING: THE BIG PICTURE

Throughout this article we've looked at social computing in terms of how social computing systems work as *systems*: they create platforms for social interaction whose results can be drawn upon by the system to add value. This is natural because we have proceeded by looking closely at examples. However, I'd like to wrap this up by looking at social computing in a subtly different way – as a type of approach to computation.

In my view, social computing is not so much about computer systems that accommodate social activity, but rather it is about systems that perform computations on information that is embedded in a social context. That is:

> "*Social computing refers to systems that support the gathering, processing and dissemination of information that is distributed across social collectives. Furthermore, the information in question is not independent of people, but rather is significant precisely because it linked to people, who are in turn associated with other people.*"

At the core of this definition is the linking of information to identity. That is, information is associated with people, and, for the purposes of social computing, the association of information with identity matters. "Identity" does *not* necessarily mean that information is associated with a particular, identifiable individual. For the purposes of social computing, identity can run the gamut from guaranteeing distinctiveness (i.e., that different pieces of information come from distinct individuals, as one would want in a plebiscite), to knowing some of the characteristics of each individual with whom information is associated (a set of book purchases by a distinct but anonymous individual), to knowing a person's real world identity.

A second element of this definition is the idea that individuals are associated with one another in social collectives. Social collectives can be teams, communities, organizations, markets, cohorts, and so forth. That is, just as information is linked to a person, so are individuals associated with one another: it matters who is associated with whom, and how and why they are related. That is notto say that individuals are necessarily linked to one another in *person to person* relationships. Individuals may be mutual strangers, and "associated" only because they happen to share some characteristic like an interest in a particular book, or in MatLab programming.

In fact, in some cases, social computing systems are predicated on the assumption that individuals will be mutually anonymous. For example, markets and auctions attract participants with shared interests, but the underlying social computing mechanisms are designed to prevent individuals from identifying one another. A market functions most effectively when the actions of individuals are independent; otherwise, individuals can collude to affect the functioning of the market to benefit themselves, as when auctions are manipulated via shilling (false bids intended to raise the final price). In short, it is precisely because the linkage between information and individuals matters that, for the purposes of some social computations, it must be suppressed.

A third element of the definition is that social computing systems have mechanisms for managing information, identity, and their interrelationships. This follows from the mention of the gathering, use and dissemination of information distributed across social collectives. Whereas an ordinary computational system need only manage information and its processing, social computing systems must also manage the social collective, which is to say that it must provide a way for individuals to have in-system identities, relate information to those identities, and manage relationships among the identities (which, as noted, can include maintaining mutual anonymity, as in the case of markets). Social computing systems can take a number of approaches to this, and the sort of social architecture it employs fundamentally shapes the nature of the system.

## 4.7 WHAT'S NEXT?

Social computing is a large area, and it is one that is growing rapidly. The examples we've looked at – Pagerank, the Amazon review mechanism, the MatLab Programming Contest, the ESP Game, and Wikipedia – just scratch the surface. New examples of social computing mechanisms and systems spring up seemingly over night.

This article has focused on 'conventional' examples of social computing. That is, they are web-based and largely draw on and appeal to educated audiences spread across the industrialized world. We are beginning to see, and will see many more, social computing systems that are designed expressly for mobile devices, that are targeted locally rather than globally, and that will include or be expressly targeted at populations in developing regions. These new domains, and the challenges they pose, will shape the further development of social computing.

Social computing is evolving with great rapidity. Designers and scholars from a wide range of disciplines – behavioral economics, computer science, game design, human-computer interaction, psychology, and sociology, to name a few – are actively studying social computing systems and applying insights gleaned from their disciplines. It is difficult to predict the future, but it seems safe to say that social computing mechanisms and systems will continue to transform the way we live, learn, work and play.

## 4.8 WHERE TO LEARN MORE

The roots of social computing stretch back decades. A good starting point for those interested in the forerunners of social computing is Howard Rhinegold's *The Virtual Community* (Rheingold 1993). For those interested in a deeper taste of history, Starr Roxanne Hiltz and Murray Turoff's *The Network Nation* (Hiltz and Turoff 1993) – originally published in the late 70's and revised in the early 90's – offer an early yet comprehensive vision (portions of which still seem remarkably prescient) of forms of social intelligence and action mediated by computer networks.

For those interested the new wave of social computing systems that have been the focus of this article, the best place to begin are books that lay out the underlying rationale for social computing by showing what can be achieved by large scale collective action. *The Wisdom of Crowds*, by James Suroweicki (Suroweicki 2004), is an excellent introduction to a wide range of examples of social computation. Arriving in the same territory from a different direction is Eric von Hippel's *Democratizing Innovation* (Hippel 2005), which examines how innovation arises from groups and communities and argues for redesigning business practices and government policies to take advantage of large scale innovation.

Those who wish a more detailed understanding of social computing systems as systems – how to design them, how to launch them, how to maintain them – are at the frontiers of knowledge. A variety of workshops and symposia focus on various 'slices' through social computing. Wikisym is a central place for research involving the use of wiki's for collaboration, and the best place to plumb the growing pool of research on Wikipedia. The Human Computation workshop (Human Computation Workshop 2011), now in its third year, examines systems like the ESP game that engage large numbers of people in performing computations and other tasks. There is currently no annual conference that covers the full range of social computing systems, although perhaps that role will be filled by the Collective Intelligence conference to be inaugurated in 2012.

**WikiSym - International Symposium on Wikis**

2011 2010 2009 2008 2007 2006 2005

## 4.9 ACKNOWLEDGEMENTS

Thanks to Robert Farrell for helpful comments on the penultimate draft of this article.

## 4.10 NOTES

1. Amazon.com is a trademark of Amazon.com, Inc.; Google and Pagerank are trademarks of Google, Inc.; eBay is a trademark of eBay, Inc.; MatLab is a trademark of The Mathworks; Twitter is a trademark of Twitter, Inc.; Wikipedia is a registered trademark of the Wikimedia Foundation.

## 4.12 COMMENTARY BY ELIZABETH CHURCHILL

How to cite this commentary in your report

### Elizabeth Churchill

Elizabeth Churchill is a Principal Research Scientist and manager of the Internet Experiences group at Yahoo! Research. She previously worked at PARC, the Palo Alto Research Center, and before that at FXPAL, Fuji Xerox's research lab based in Silicon Valley where she led the Social Computing Group. Elizabeth has a BSc in Experimental Psychology, an MSc in Knowledge Based Systems, both fr...

Elizabeth Churchill
Elizabeth Churchill is a member of The Interaction Design Foundation

### 4.12.1 That was then, this is now and what's next - Tom Erickson talks about 'social computing'

IBM's Tom Erickson has been an active researcher in the field of *social computing* since its inception in the late 1990's. This set of interviews and his chapter that accompanies them is a lovely exposition on the research, design and development activities that make up social computing.

This discussion is well grounded in theory, in practice and in a rich set of analyses of existing systems. Tom offers his perspective on the coverage and concerns of the area of social computing with recourse to existing technologies: those that he has been directly engaged in designing and building (e.g., the Babble system developed and used at IBM [1]); systems that are based in research from university groups (e.g., the Grouplens Research group's fabulous example of collective, 'crowd' knowledge in the Cyclopath bike route finder [2]); and some everyday social resources and services that we all know and love (e.g., Amazon, Wikipedia, eBay).

Tom's exposition engages with all of my favourite "C words"— cooperation, collaboration, communication, conversation, competition, congregation, collective and what seems to be a relatively recently invented word 'coopetition' meaning cooperative competition. These words describe different forms of social organization and orientation for interaction with others. They define the ways in which social mechanisms and social skills are used in conducting ourselves with others to achieve personal and collective goals. Two more "C" words can be added to this collection and are discussed in essence — these are 'coercion' and 'conned'. Utopian perspectives focus on the co-development of social arrangements and the establishment of consent; the darker, dystopian perspective is that social skills and manipulative mechanisms may be used to drive non-consensual outcomes through coercion and/or through being "conned". As we are all aware, the truth about social computing systems is that they support the full panoply of human social engagement, from angelic to demonic. To illustrate how these different forms of social engagement are supported and enacted, Tom covers a number of social technologies, from low-social-touch systems that aggregate people's viewpoints in reviews and recommendations (such as the reviews and recommendations on a shopping site like Amazon) to sites where mediated interactions are richer but highly structured (such as auction sites like eBay) to high-social-touch, relatively technologically unstructured and conversational online places that support text chat (e.g., chat spaces and instant messenger), and avatar-embodied interactions in online virtual world settings like (e.g., Second Life). One could of course add inmediated, embodied audio/video interactions (e.g., Chatroulette).

As Tom speaks about the area of social computing, we are reminded that the word "social" is a tricky adjective. As I have written about elsewhere, the term can mean many different things to different people [1]. Wisely, Tom is careful to define what he means when he says 'social computing' — he defines social computing to be any situation in which people are involved in the computation. With recommendations on sites like Amazon these computations are often made without people's active involvement in the computation of the recommendation whereas in distributed collaboration and expertise sharing settings (like Slashdot for example), people may be actively collaborating. In articulating his definition of 'social computing', Tom distinguishes between social media and social computing — in social media no explicit computation is taking place. However, I would argue that many of the more general observations he makes are as appropriate for social media systems as they are for social *computing* systems.

In discussing the design of social computing technologies, considerable weight is given, as it should be, to the design and development process itself. Tom's favored approach is to observe human sociality as it takes place in face-to-face, embodied settings and to draw analogies to what occurs and/or could be effectively replicated in the online world. This also requires that we get off our sofas and out of our offices and open our eyes to how social interactions take place beyond our own well worn-paths, away from our familiar settings. Through illustrative examples Tom reminds us that in designing, it is easy to become myopic, to forget that the world is not necessarily configured for everyone the way it is for us. Tom's design sensibilities, in this regard, are close to those of participatory design [5]. Personally, I am deeply sympathetic to the 'get-off-your-sofa approach to design and to evaluation, and strongly believe that *design/develop*, *launch*, *watch*, *ask*, *iterate* and *repeat (*a.k.a. DELAWAIR) with an open mind are the keys to successful design in social computing. Premature design commitments and ossified interaction/infrastructures usually lead to short-lived social systems. Indeed, Tom makes clear that his development sensibilities tend to the building of mutable tools that are tested early, iteratively designed and that can be appropriated by users to their own ends — which, notably, may have nothing in common with the original design intentions.

The idea of conducting field investigations that open our eyes to differences in ways of thinking and different norms for social action is not new, but it is easy to forget to look out for how our technologies are being adopted, adapted and indeed appropriated. Tom reminds us to move beyond simple characterizations of other perspectives and to field our technologies with a view to being surprised. Indeed, he suggests if we are *not* surprised, perhaps we are not designing well enough. The humility of this approach is very appealing to me.

There's a lot of rich detail in this chapter and these wonderful video interviews, too much to do justice to in a short commentary. My mind is a whirl with ideas and thoughts and reactions and inspirations. Therefore, I will pick up on a couple of areas of design in social computing are mentioned and that are ripe for further investigation, especially for those with critical HCI tendencies, inclusive intentions and experience design leanings. These are not new areas *per se* but the world is changing. Many of us started our careers in social computing looking at the use of such technologies within fairly prescribed social groups and organizations —designing for people who were already grouped in some way, who worked within an organization or who already had a strong sense of homophily or like-mindedness. However, the world is changing and just being there, on the site, is no longer a guarantee of homophily. Nor are there, necessarily, societal or organizational macro structures governing or shaping adoption and use. We are living in a world of increasingly powerful consumer technologies, an increase in global telecommunications networks, and, as a result, expanding access to applications that enable connections with known and unknown others. There are enormous social, political and economic ramifications of these changes of course [e.g. for excellent reading on these broader implications 6, 7, 8]. However, for the purposes of this commentary, there are a couple of areas of interest that remain close to the experiences of individuals but which deserve more research investigation for how sociality and social computing are changing (or not); we may also consider in this how upcoming technological innovations will change the field of social computing experience itself. A couple of my favourite areas that Tom alludes to are: (1) emotion, (2) embodiment, (3) literacy and access, construed in the broadest sense, and finally, (4) data and instrumentation. I'll quickly talk about each in turn.

### 4.12.2 Emotion

We know that the connections people make with others through these systems are not just useful and transactional but also emotional; consider the 'high' of 'winning' an auction, the emotional roller-coaster of being on a dating site, the thrill of seeing a comment on a photo one has posted, the warmth one feels when a loved one sends a message through Facebook, the satisfaction one gets in seeing one's points go up in Slashdot because of one's recognized expertise and the sense of anger, fear and bewilderment when one is betrayed, assaulted or abused online. Happiness, sadness, anger are all present in our interactions in social computing systems. Emotions are not just in-the-moment and ephemeral; emotions in the now matter for my actions in the future. Emotions underlie action and abdication; emotions foster avoidance or participation. Emotions orient us toward places and spaces where we feel good, and propel us away from places where we feel violated. Like pleasures, violations can come in many forms: threatened by the service providers (what are you doing with my data?); threatened by others on the site (why are you so obnoxious to me?) and by the tenor of the site itself (goodness, this is in poor taste). There are also cultural differences that need to be taken account of;

anthropologists talk of differing emotional styles, and discuss norms of emotional response—how, when and where emotions like joy, anger, pride, shame, embarrassment are expressed. Emotions are important. They matter for individual and social computation. We know emotions drive participation or avoidance, but they also underlie cognition. António Damásio argues convincingly that emotions are involved in decision-making and are the basis for social cognition. Indeed he claims they are the cornerstone of consciousness. [9] How can we understand the ways in which social computing technologies are altering our emotional landscape? In turn, how is social computing shaped by the kinds of emotions that are expressed onsite and/or (in)expressible through the technologies themselves? What. If any, issues arise from cultural differences in the expression of human emotion? What are the effects of individual and collective emotion on collective cognition?

### 4.12.3 Embodiment

Many of Tom's examples relate the conduct of people in physical places. Urban spaces, as he points out, are changing with the introduction of embedded sensors, within which category I include the proliferation of surveillance monitors. Embodied social computing and social computing of space and place are necessarily locative. The social computation in and of urban, social spaces is changing as we check-in to location-based social networking sites and track others individually and in the aggregate. Certainly, researchers interested in 'reality mining' are already tracking us, and applications allow us to easily see what are the most popular, most trafficked parts of town, places for us to avoid or navigate to, as we see fit. Perhaps the Cyclopath Geowiki project is a good example of how social computation is altering human ambulation through physical places as a result of the introduction of new forms of intentionally submitted and tacitly collected data. More mundanely, our behaviours in urban spaces with respect to each other are changing as we engage through our devices with others who are not physical present. Who has not collided with someone whose eyes and ears were directed firmly at a communication device as they walked blithely and ignorantly forward? Do these changes in the navigation of physical and social space and place have implications for social computation?

### 4.12.4 Literacy and access

Tom makes a nice point about symmetry and asymetries in social computing. Who does and does not have access and through what tools can create barriers to social participation. Literacy and access is central to the inclusive project of social computing. But exclusion is not simply a problem because it is a matter of not taking active part. Those who are not active are also not represented in models of sociality that are based on data from networked systems; the early intentions of anthropologists interested in human social networks was to understand human relationships, not just those that were easily available in accessible data sets [see 4]. This brings me to my final point of interest, again something to which Tom's analyses point us.

### 4.12.5 Data analytics

As social scientists, designers, developers and business people we are very keen on measurement, on metrics for success and instrumented systems through which we can interrogate social activities and develop patterns on the basis of which we make claims about the fundamentals of human sociality. When social computing systems were only used in small groups in work contexts, the methods we used to triangulate our understanding were plentiful even if they were not as easily proceduralized as systematized datamining. Usage data analysis was combined with survey and/or interviews and/or an observational study. However, with the use of social computing technologies on a large scale—on the scales of hundreds of thousands and millions to many millions of people—scale bites us. It's the seduction of large numbers and the delight of aggregates. We seem to have lost the art of designing effective instrumentation in the large to satisfy the rigors of good science and build meaningful models of systems-in-use. Typically, with the exception of a few exemplar systems, we are not instrumenting for experience analytics. 'Big data' may be everyone's favourite fetish right now, but we need to more deeply understand how a sample is drawn, how many populations are represented on a social computing platform, from what population it derives, and the extent to which that population is representative of what and who we care about. If we can't understand such basic questions, we have a seductive substrate for description and the fodder for hyperbolic claims. But it is not clear to me we really understand much more about human sociality that I could not have told you with a peek at Maslow's hierarchy of needs [10].

Looking at this more positively, we are in a world of amazing opportunity. From a science, design, development, statistics and business of social computing perspective, we live in exciting times. That said, Tom touches on ethics throughout his contributions. His comments remind us that, in the surging world of Internet scale social computing in the midst of big numbers and exciting technological capabilities, it behooves us to look to the people who are using our technologies, to how they compute collectively with and through our tools, and to how our computations of their activities reflect *them*.

### 4.12.6 Citations/references


- [1] Erickson, T., Smith, D. N., Kellogg, W. A., Laff, M. R., Richards, J. T., and Bradner, E. "Socially Translucent Systems: Social Proxies, Persistent Conversation, and the Design of 'Babble.'" In Human Factors in Computing Systems: The Proceedings of CHI '99. ACM Press, 1999.
- [2] For more on the Cyclopath Geowiki project from the Grouplens Research group in Minnesota see cyclopath.org

- [3] Elizabeth F. Churchill| Socializing at Cross Purposes, *interactions,* Volume 17 Issue 1, 2010. New York: ACM Press
- [4] Elizabeth F. Churchill The (Anti) Social Net, *interactions*, Volume 17, Issue 5. 2010/ New York: ACM Press
- [5] Bødker, K., Kensing, F., and Simonsen, J. (2004). Participatory IT design: Designing for business and workplace realities. Cambridge, MA, USA: MIT Press.
- [6] Manuel Castells, The Information Age trilogy: Castells, Manuel (1996, second edition, 2000). The Rise of the Network Society, The Information Age: Economy, Society and Culture Vol. I. Cambridge, MA; Oxford, UK: Blackwell. ISBN 978-0631221401. Castells, Manuel (1997, second edition, 2004). The Power of Identity, The Information Age: Economy, Society and Culture Vol. II. Cambridge, MA; Oxford, UK: Blackwell. ISBN 978-1405107136. and Castells, Manuel (1998, second edition, 2000). End of Millennium, The Information Age: Economy, Society and Culture Vol. III. Cambridge, MA; Oxford, UK: Blackwell. ISBN 978-0631221395.
- [7] Manuel Castells, Communication Power, Oxford University Press, 2009
- [8] Jochai Benkler, The Penguin and the Leviathan: How Cooperation Triumphs over Self-Interest. New York: Random House, 2011
- [9] António DamásioDescartes' Error: Emotion, Reason, and the Human Brain, Putnam, 1994; revised Penguin edition, 2005
- [10] A.H. Maslow, A Theory of Human Motivation, Psychological Review 50(4) (1943):370-96.

## 4.13 COMMENTARY BY DAVID W. MCDONALD

How to cite this commentary in your report

### David W. McDonald

© *David W. McDonald*

David W. McDonald is Associate Professor at the University of Washington. He has an MS (1992) in Computer Science, California State University Hayward; and an MS (1995) and PhD (2000) in Information and Computer Science. His research interests include Computer-supported cooperative work (CSCW), human-computer interaction (HCI), collaborative systems design, software architecture, softwar...

David W. McDonald
David W. McDonald is a member of The Interaction Design Foundation

Commentary coming soon. David is almost done. Please stay tuned.

## 4.14 COMMENTARY BY ANDREA FORTE

How to cite this commentary in your report

### Andrea Forte

© *Andrea Forte*

I'm an assistant professor in the iSchool (College of Information Science and Technology) at Drexel University in Philadelphia. I received my PhD from the College of Computing at Georgia Institute of Technology in human-centered computing with a focus on social computing and learning sciences. My MLIS is from the School of Information at University of Texas at Austin. I'm intereste...

Andrea Forte
Andrea Forte is a member of The Interaction Design Foundation

It is hard to imagine a topic that exposes the power of good design with more salience than social computing. Creating systems in which people debate, teach, hate, play, bully and love one another is a phenomenal line of work.

Central to social computing is the idea that traces of computationally mediated activities can be exposed to users to help them make choices about how to behave. Furthermore, these traces can be aggregated or otherwise computationally manipulated and fed back into socio-technical systems to form a feedback loop. But what kinds of information should be surfaced? How can it be usefully manipulated? And what kind of influence will that information have on people's behavior? These questions are the bread and butter of social computing design and research.

Tom Erickson begins this chapter with a simple but powerful statement, "As humans we are fundamentally social creatures." He goes on to describe the seemingly effortless ways that average people use and create social information in their everyday lives. He talks about the challenge of designing to support interaction by reproducing or compensating for the lack of cues from the physical world. One particularly memorable proposition is that computational systems must support deception in order to support socially graceful interaction among humans.

But, in order to lie, you have to know how.

The "natural" ways that humans convey emotion through demeanor, construct identities by selecting words and dress, and reassure each other through feigned attention and other "little white lies" are skills we learn and refine throughout our lives. These social skills require different knowledge and skills in a computationally mediated world.

An example. On the social music site Last.fm, members may publish their music libraries and playlists and a log of all the music they listen to, updated in real time. To project the image of a Lady Gaga fan on Last.fm, it is not sufficient to simply state that one listens to Lady Gaga, one must actually listen to her so the software can record that activity and expose it. Others can observe the record of your listening experiences and use it as a recommendation of what to listen to (or avoid). In order to project a desired image, fans have been known to manipulate tracks so that they appear to play several times in a minute, artificially boosting their personal playcount of newly released or beloved songs and artists and influencing sitewide statistics. To do so requires proficiency with digital audio.

Computational deception requires computational literacies.

One might argue that the best designs will prove to be intuitive as users perceive and respond to social cues that map closely to their face-to-face analogs. But beyond the enticing design problems raised by the challenge of enabling humans to engage with one another "naturally" in mediated environments, there is the question of what kinds of things people might do in mediated environments that they were never able to do before. What kinds of *new* behaviors might socio-computational systems exhibit? What new manifestations of power might people—users and designers alike—wield over one another? And what new competencies will these new possibilities require of people?

Tom Erickson builds a compelling vision of new possibilities for mediated social interaction on a simple foundation: Humans are fundamentally social. We depend on social information to carry out the simplest tasks. We generate social information as a byproduct of every activity we carry out. What to do with that information is one of the most enticing, meaningful and complex design problems of the coming decades.

## 4.11 BEHIND THE SCENES

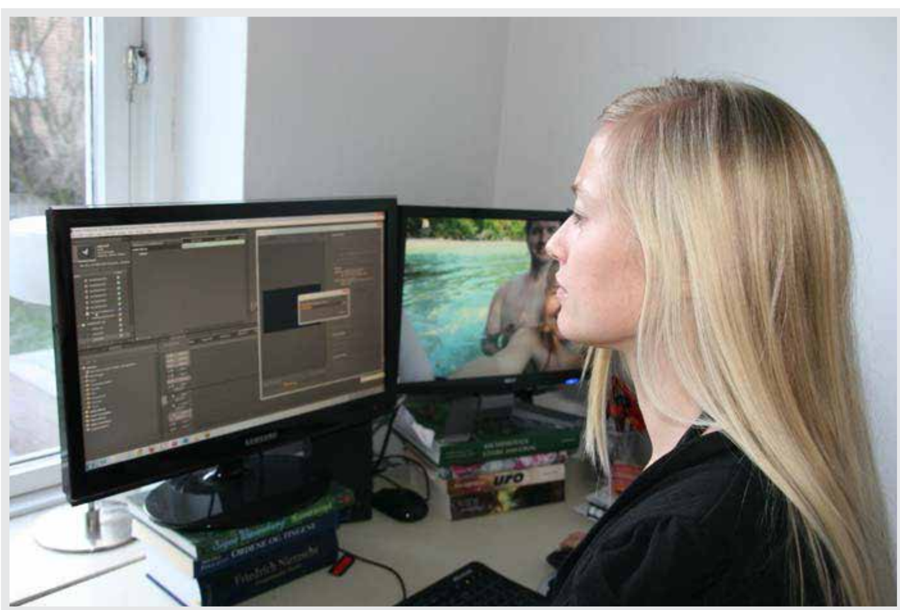
Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.

Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.

Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.

Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.

Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.



Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.

## 4.15 REFERENCES

*Ahn, Luis von and Dabbish, Laura (2004): Labeling images with a computer game. In: Dykstra-Erickson, Elizabeth and Tscheligi, Manfred (eds.) Proceedings of ACM CHI 2004 Conference on Human Factors in Computing Systems April 24-29, 2004, Vienna, Austria. pp. 319-326*

*Ahn, Luis von and Dabbish, Laura (2008): Designing games with a purpose. In Communications of the ACM, 51 (8) pp. 58-67*

*Butler, Brian, Joyce, Elisabeth and Pike, Jacqueline (2008): Don't look now, but we've created a bureaucracy: the nature and roles of policies and rules in wikipedia. In: Proceedings of ACM CHI 2008 Conference on Human Factors in Computing Systems April 5-10, 2008. pp. 1101-1110*

*Erickson, Thomas (1996): The world wide web as social hypertext. In Communications of the ACM, 31 (1) pp. 15-17*

*Giles, Jim (2005): Internet encyclopaedias go head to head. In Nature, 438 pp. 900-901*

*Guardian, The (2011). Investigate Your MP's Expenses. Retrieved 19 June 2011 from The Guardian - guardian.co.uk:*

*Gulley, Ned (2004): In praise of tweaking: a wiki-like programming contest. In Interactions, 11 (3) pp. 18-23*

*Hiltz, Starr Roxanne and Turoff, Murray (1993): The Network Nation: Human Communication via Computer. MIT Press*

*Hippel, Eric von (2005): Democratizing Innovation. The MIT Press*

*Human Computation Workshop (2011). The Human Computation Workshop. Retrieved 22 July 2011 from Human Computation Workshop: http://www.humancomputation.com/*

*Kittur, Aniket and Kraut, Robert E. (2008): Harnessing the wisdom of crowds in wikipedia: quality through coordination. In: Proceedings of ACM CSCW08 Conference on Computer-Supported Cooperative Work 2008. pp. 37-46*

*Kittur, Aniket, Suh, Bongwon, Pendleton, Bryan A. and Chi, Ed H. (2007): He says, she says: conflict and coordination in Wikipedia. In: Proceedings of ACM CHI 2007 Conference on Human Factors in Computing Systems 2007. pp. 453-462*

*Licklider, J. C. R. and Taylor, Robert W. (1968): The Computer as a Communication Device. In Science and Technology, pp. 21-31*

*MatLab Central (2010). MatLab Central: MatLab Programming Contest. Retrieved 19 June 2010 from MatLab Central: http://www.mathworks.com/matlabcentral/contest/*

*MatLab Central (2011). MatLab Central: Molecule Contest Analysis. Retrieved 19 June 2011 from MatLab Central:*

Panciera, Katherine, Halfaker, Aaron and Terveen, Loren (2009): Wikipedians are born, not made: a study of power editors on Wikipedia. In: GROUP09 - International Conference on Supporting Group Work 2009. pp. 51-60

Priedhorsky, Reid, Chen, Jilin, Lam, Shyong (Tony) K., Panciera, Katherine, Terveen, Loren and Riedl, John (2007): Creating, destroying, and restoring value in wikipedia. In: GROUP07: International Conference on Supporting Group Work 2007. pp. 259-268

Raddick, M. Jordan, Bracey, Georgia, Gay, Pamela L., Lintott, Chris J., Murray, Phil, Schawinski, Kevin, Szalay, Alexander S. and vandenBerg, Jan (2010): Galaxy Zoo: Exploring the Motivations of Citizen Science Volunteers. In Astronomy Education Review, 9 (1)

Ramo, Simon (1961): The Scientific Extension of the Human Intellect. In Computers and Automation, 10 (2) pp. 9-12

Rheingold, Howard (1993): The Virtual Community: Homesteading on the Electronic Frontier. Reading, MA, The MIT Press

Surowiecki, James (2004): The Wisdom of Crowds. Anchor

Viégas, Fernanda B., Wattenberg, Martin and Dave, Kushal (2004): Studying cooperation and conflict between authors with history flow visualizations. In: Dykstra-Erickson, Elizabeth and Tscheligi, Manfred (eds.) Proceedings of ACM CHI 2004 Conference on Human Factors in Computing Systems April 24-29, 2004, Vienna, Austria. pp. 575-582

Wilkinson, Dennis M. and Huberman, Bernardo A. (2007): Cooperation and quality in wikipedia. In: Proceedings of the 2007 International Symposium on Wikis 2007. pp. 157-164

Woolley, David R. (1994). PLATO: The emergence of online community. Retrieved 19 June 2011 from http://thinkofit.com/plato/dwplato.htm

Zoo, Galaxy (2011). Galaxyzoo.org. Retrieved 19 June 2011 from Galaxy Zoo:

# ABOUT THE AUTHOR

## Thomas Erickson

© *Thomas Erickson*

HAS ALSO PUBLISHED UNDER THE NAME OF:
"Thomas D. Erickson"

PERSONAL HOMEPAGE:
http://www.visi.com/~snowfall/

CURRENT PLACE OF EMPLOYMENT:
IBM Watson Research Center

I'm an interaction designer and researcher in the Social Computing Group at IBM's Watson Labs in New York to which I telecommute from my home in Minneapolis. My research focuses on designing systems that enable groups of people to interact coherently and productively: originally focused on online systems, the scope of my work has expanded to include real world environments ranging from rooms to cities. More generally, I am interested in topics such as genre theory, pattern languages, urban design, real and virtual communities, and the sociology of human-human interaction, all of which inform my approach to systems design. I've been at IBM since June '97; before that I spent 9 years at Apple, and before that 5 years in a now-defunct startup that competed with another startup called Lotus.

## YOUR NOTES AND THOUGHTS ON CHAPTER 4

Record your notes and thoughts on this chapter. If you want to share these thoughts with others online, go to the bottom of the page at:
http://www.interaction-design.org/encyclopedia/social_computing.html

**NOTES:**

# CHAPTER 5

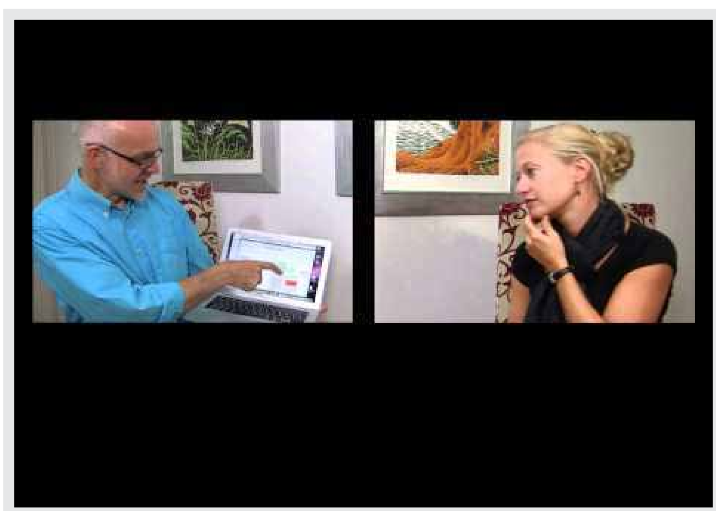
# Visual Representation

***by Alan Blackwell.***

How can you design computer displays that are as meaningful as possible to human viewers? Answering this question requires understanding of visual representation - the principles by which markings on a surface are made and interpreted. The analysis in this article addresses the most important principles of visual representation for screen design, introduced with examples from the early history of graphical user interfaces. In most cases, these principles have been developed and elaborated within whole fields of study and professional skill - typography, cartography, engineering and architectural draughting, art criticism and semiotics. Improving on the current conventions requires serious skill and understanding. Nevertheless, interaction designers should be able, when necessary, to invent new visual representations.

VIDEO 5.1: Introduction to Visual Representation by Alan Blackwell.





VIDEO 5.2: Alan Blackwell on applying theories of Visual Representation.

## 5.1 TYPOGRAPHY AND TEXT

For many years, computer displays resembled paper documents. This does not mean that they were simplistic or unreasonably constrained. On the contrary, most aspects of modern industrial society have been successfully achieved using the representational conventions of paper, so those conventions seem to be powerful ones. Information on paper can be structured using tabulated columns, alignment, indentation and emphasis, borders and shading. All of those were incorporated into computer text displays. Interaction conventions, however, were restricted to operations of the typewriter rather than the pencil. Each character typed would appear at a specific location. Locations could be constrained, like filling boxes on a paper form. And shortcut command keys could be defined using onscreen labels or paper overlays. It is not text itself, but keyboard interaction with text that is limited and frustrating compared to what we can do with paper (Sellen and Harper 2001).

But despite the constraints on keyboard interaction, most information on computer screens is still represented as text. Conventions of typography and graphic design help us to interpret that text as if it were on a page, and human readers benefit from many centuries of refinement in text document design. Text itself, including many writing systems as well as specialised notations such as algebra, is a visual representation that has its own research and educational literature. Documents that contain a mix of bordered or coloured regions containing pictures, text and diagrammatic elements can be interpreted according to the conventions of magazine design, poster advertising, form design, textbooks and encyclopaedias. Designers of screen representations should take care to properly apply the specialist knowledge of those graphic and typographic professions. Position on the page, use of typographic grids, and genre-specific illustrative conventions should all be taken into account.

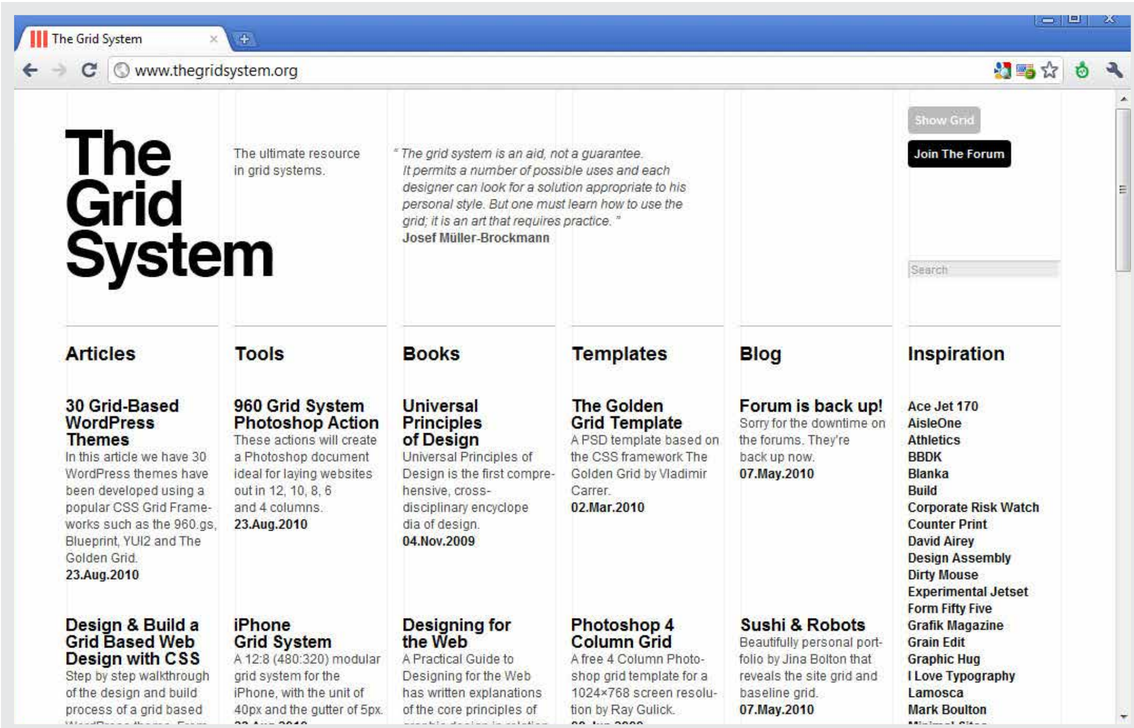



Figure 5.1: Contemporary example from the grid system website.

*Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.*

$$i\hbar\frac{\partial}{\partial t}\Psi = -\frac{\hbar^2}{2m}\nabla^2\Psi + V\Psi$$

Figure 5.2: Example of a symbolic algebra expression (the single particle solution to Schrodinger's equation).

*Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.*

# A Table of the Funerals
## IN THE
## Several Pariſhes within the Bills of Mortality of the City of *LONDON*,
### For the Year 1665.

| | No. of Funerals | Plag. | | No. of Funerals | Plag. | | No. of Funerals | Plag. |
|---|---|---|---|---|---|---|---|---|
| ST. Albans Woodſtreet | 200 | 121 | St. George Botolph-lane | 41 | 27 | St. Martins Ludgate | 196 | 128 |
| St. Alhallows Barkin | 514 | 330 | St. Gregories by Pauls | 376 | 232 | St. Martins Orgars | 110 | 71 |
| t. Alhallows Breadſtreet | 35 | 16 | St. Helens | 108 | 75 | St. Martins Outwich | 60 | 34 |
| t. Alhallows the Great | 455 | 426 | St. James Dukes-place | 262 | 190 | St. Martins Vintrey | 417 | 349 |
| t. Alhallows Honey-lane | 10 | 5 | St. James Garlickhith | 189 | 118 | St. Matthew Friday-ſtreet | 24 | 6 |
| t. Alhallows the Leſs | 239 | 175 | St. John Baptiſt Walbrook | 138 | 83 | St. Maudlins Milk-ſtreet | 44 | 22 |
| t. Alhallows Lombardſtreet | 90 | 62 | St. John Evangeliſt | 9 | | St. Maudlins Old Fiſh-ſtreet | 176 | 121 |
| t. Alhallows Staining | 185 | 112 | St. John Zachary | 85 | 54 | St. Michael Baſſiſhaw | 253 | 164 |
| t. Alhallows the Wall | 500 | 350 | St. Katherine Coleman-ſtreet | 299 | 213 | St. Michael Cornhil | 104 | 52 |
| t. Alphage | 271 | 115 | St. Katherine Cree-church | 335 | 201 | St. Michael Crooked-lane | 179 | 133 |
| t. Andrew Hubbard | 71 | 25 | St. Lawrence Jewry | 94 | 48 | St. Michael Queenhith | 203 | 122 |
| t. Andrew Underſhaft | 274 | 189 | St. Lawrence Pountney | 214 | 140 | St. Michael Quern | 44 | 18 |
| t. Andrew Wardrobe | 476 | 308 | St. Leonard Eaſtcheap | 42 | 27 | St. Michael Royal | 152 | 116 |
| t. Anne Alderſgate | 282 | 197 | St. Leonard Foſter-lane | 335 | 255 | St. Michael Woodſtreet | 122 | 62 |
| t. Anne Black-Friers | 652 | 467 | St. Magnus | 103 | 30 | St. Mildred Bread-ſtreet | 59 | 26 |
| t. Antholins | 58 | 35 | St. Margaret Lothbury | 100 | 66 | St. Mildred Poultrey | 68 | 45 |
| t. Auſtins | 43 | 20 | St. Margaret Moſes | 38 | 25 | St. Nicholas Acons | 46 | 28 |
| t. Bartholomew Exchange | 73 | 51 | St. Margaret New Fiſhſtreet | 114 | 65 | St. Nicholas Coleabby | 125 | 91 |
| t. Bennet Finch | 47 | 22 | St. Margaret Pattons | 49 | 24 | St. Nicholas Olaves | 90 | 62 |
| t. Bennet Grace-church | 57 | 41 | Sr. Mary Abchurch | 99 | 54 | St. Olaves Hart-ſtreet | 237 | 160 |
| t. Bennet Pauls Wharf | 355 | 172 | St. Mary Aldermanbury | 181 | 109 | St. Olaves Jewry | 54 | 32 |
| t. Bennet Sherehog | 11 | 1 | St. Mary Aldermary | 105 | 75 | St. Olaves Silver-ſtreet | 250 | 132 |
| St. Botolph Billinſgate | 83 | 50 | St. Mary-le-Bow | 64 | 36 | St. Pancras Soper-lane | 30 | 15 |
| Chriſt Church | 653 | 46- | St. Mary Bothaw | 55 | 30 | St. Peters Cheap | 61 | 35 |
| Sr. Chriſtophers | 60 | 47 | St. Mary Colechurch | 17 | 6 | St. Peters Cornhil | 136 | 76 |
| St. Clements Eaſtcheap | 38 | 20 | St. Mary Hill | 94 | 64 | St. Peters Pauls Wharf | 114 | 86 |
| St. Dionys Back-church | 78 | 27 | St. Mary Mounthaw | 56 | 37 | St. Peters Poor | 79 | 47 |
| St. Dunſtans in the Eaſt | 265 | 150 | St. Mary Somerſet | 342 | 262 | St. Stephens Coleman-ſtreet | 560 | 391 |
| St. Edmunds Lombardſtreet | 70 | 30 | St. Mary Stainings | 47 | 27 | St. Stephens Walbrook | 34 | 17 |
| St. Ethelborough | 195 | 106 | St. Mary Woolchurch | 65 | 33 | St. Swithins | 93 | 56 |
| St. Faiths | 104 | 70 | St. Mary Woolnoth | 75 | 38 | St. Thomas Apoſtle | 163 | 110 |
| St. Foſters | 144 | 105 | St. Martins Ironmonger-lane | 21 | 11 | Trinity Pariſh | 115 | 79 |
| St. Gabriel Fenchurch | 69 | 35 | | | | | | |

*In the 97 Pariſhes within the Walls, Total of the Funerals* 15207; *Died of the Plague* 9887.

Figure 5.3: Table layout of funerals from the plague in London in 1665.

Figure 5.4: Tabular layout of the first page of the Gutenberg Bible: Volume 1, Old Testament, Epistle of St. Jerome. The Gutenberg Bible was printed by Johannes Gutenberg, in Mainz, Germany in the 1450s.

### 5.1.1 Summary

Most screen-based information is interpreted according to textual and typographic conventions, in which graphical elements are arranged within a visual grid, occasionally divided or contained with ruled and coloured borders. Where to learn more:

- thegridsystem.org
- Resnick, Elizabeth (2003): *Design for Communication: Conceptual Graphic Design Basics*. Wiley

## 5.2 MAPS AND GRAPHS

The computer has, however, also acquired a specialised visual vocabulary and conventions. Before the text-based computer terminal (or 'glass teletype') became ubiquitous, cathode ray tube displays were already used to display oscilloscope waves and radar echoes. Both could be easily interpreted because of their correspondence to existing paper conventions. An oscilloscope uses a horizontal time axis to trace variation of a quantity over time, as pioneered by William Playfair in his 1786 charts of the British economy. A radar screen shows direction and distance of objects from a central reference point, just as the Hereford Mappa Mundi of 1300 organised places according to their approximate direction and distance from Jerusalem. Many visual displays on computers continue to use these ancient but powerful inventions - the map and the graph. In particular, the first truly large software project, the SAGE air defense system, set out to present data in the form of an augmented radar screen - an abstract map, on which symbols and text could be overlaid. The first graphics computer, the Lincoln Laboratory Whirlwind, was created to show maps, not text.

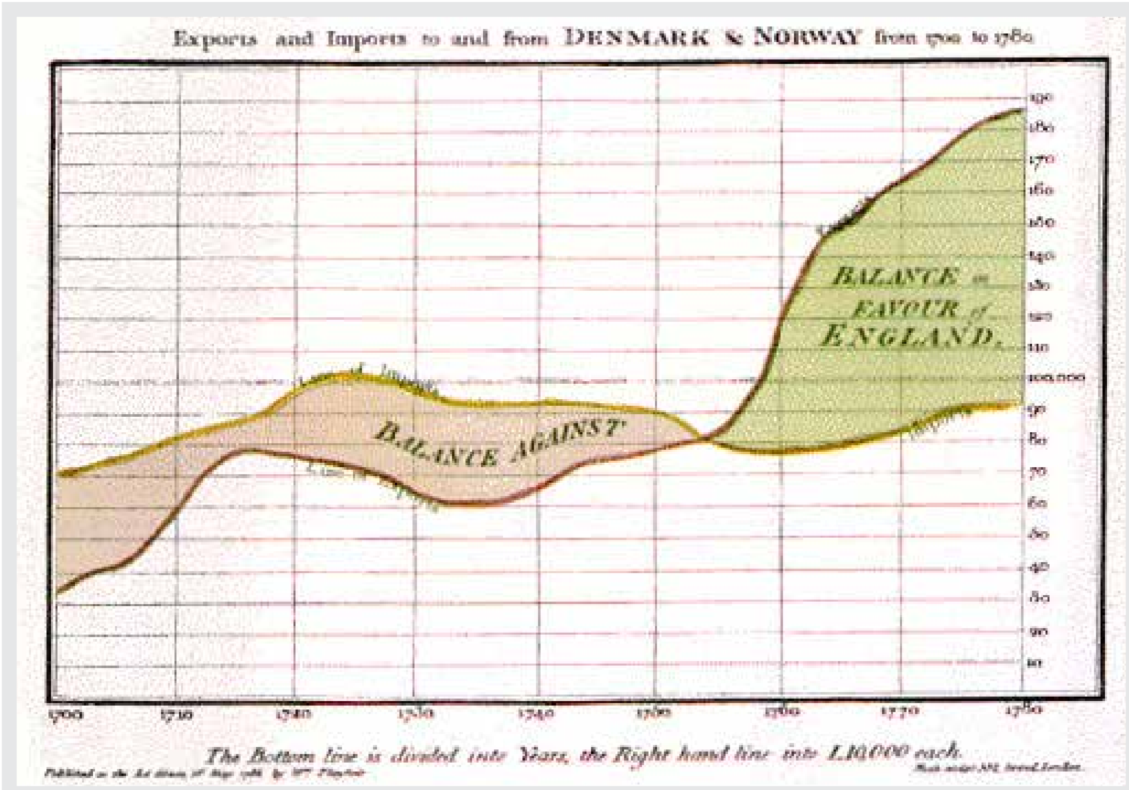



Figure 5.5: The technique invented by William Playfair, for visual representation of time series data.

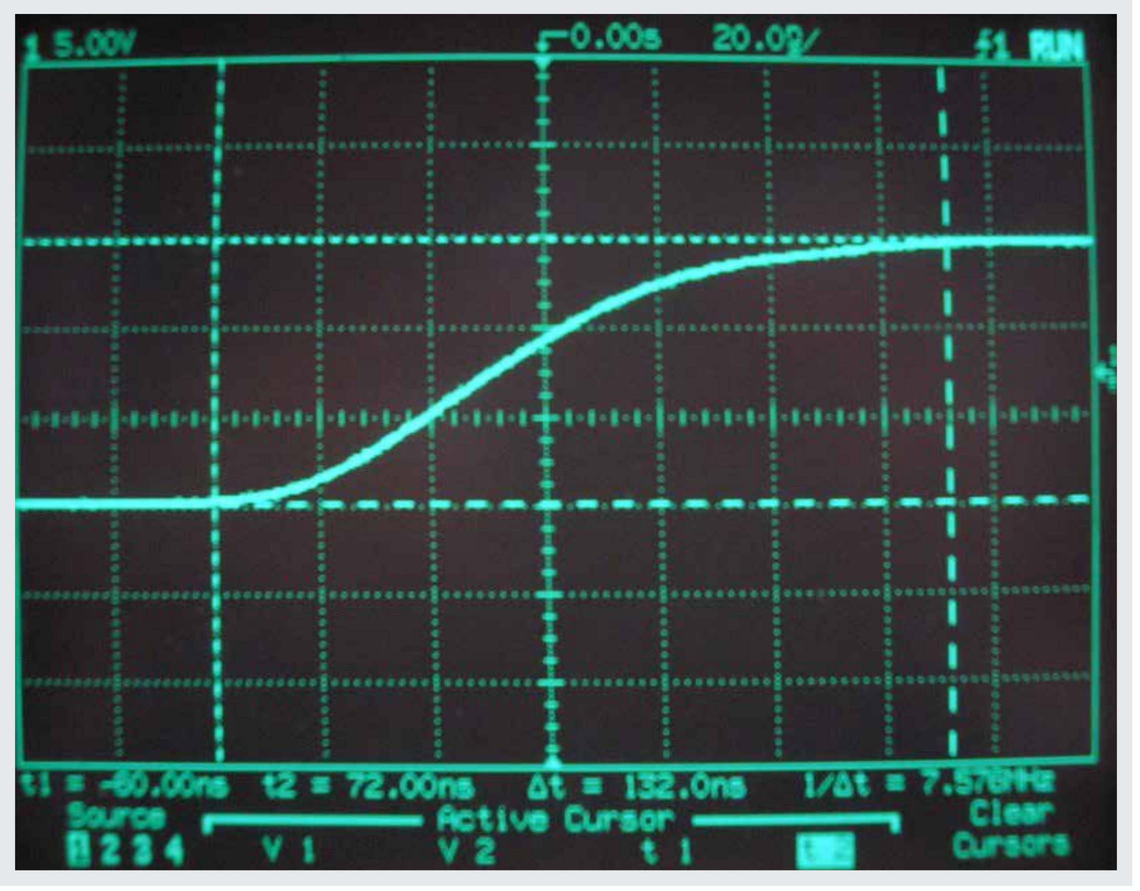



FIGURE 5.6: Time series data as shown on an oscilloscope screen.

*Courtesy of Premek. V.. Copyright: pd (Public Domain (information that is common property and contains no original authorship)).*

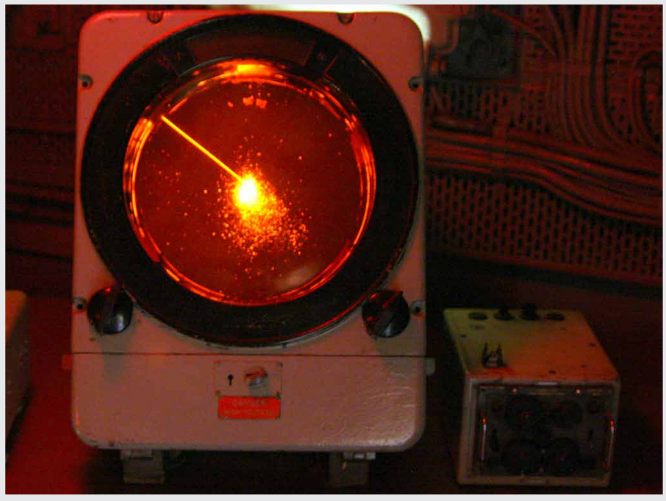


Figure 5.7: Early radar screen from HMS Belfast built in 1936.

*Courtesy of Remi Kaupp. Copyright: CC-Att-SA (Creative Commons Attribution-ShareAlike 3.0 Unported).*

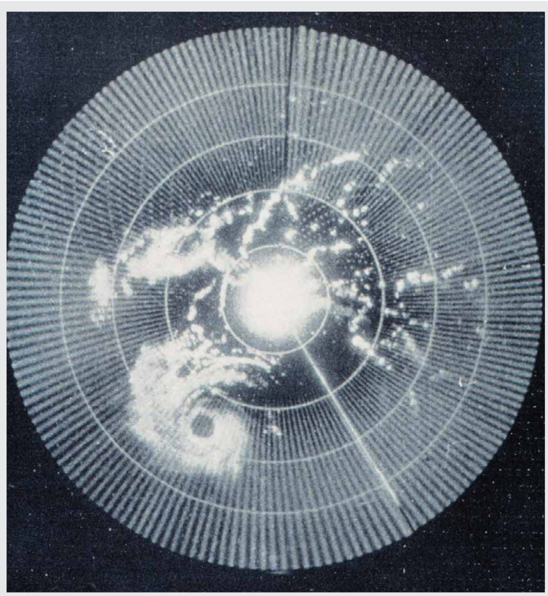


Figure 5.8: Early weather radar - Hurricane Abby approaching the coast of British Honduras in 1960.

*Courtesy of NOAA's National Weather Service. Copyright: pd (Public Domain (information that is common property and contains no original authorship)).*

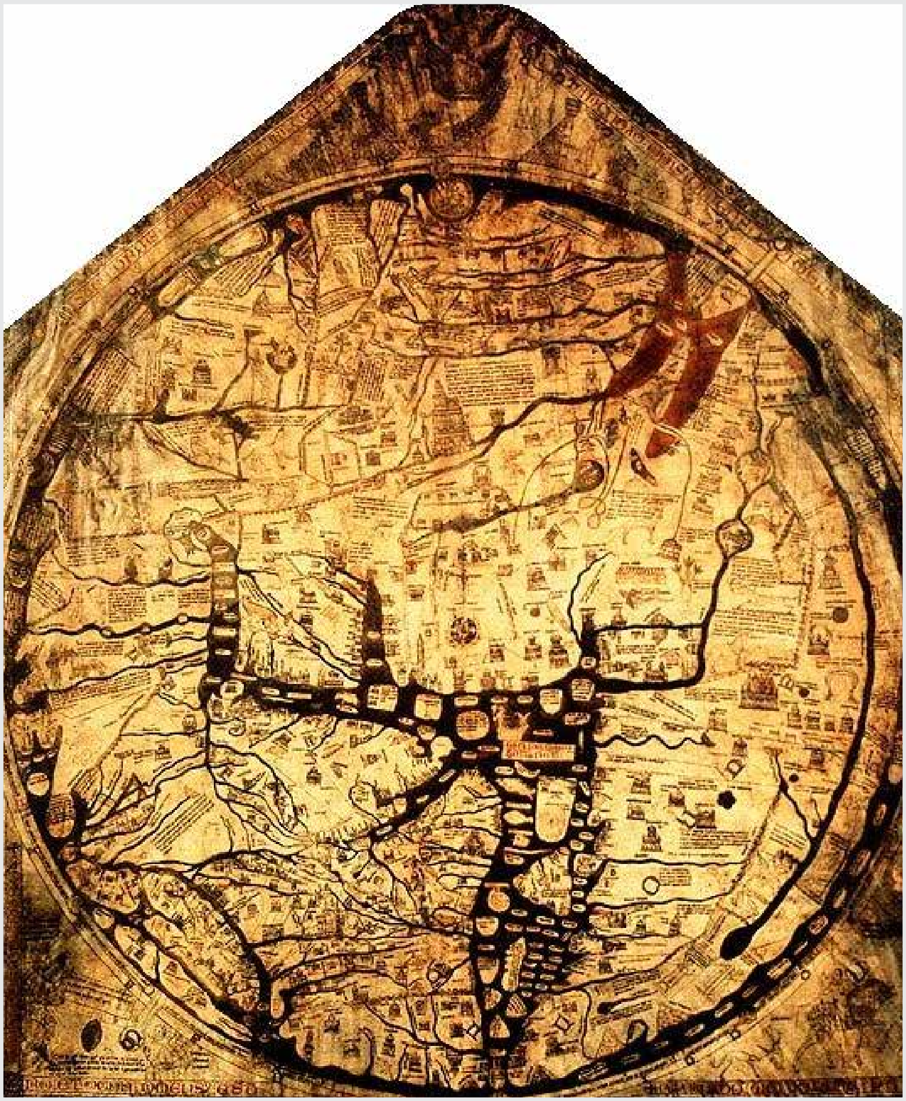


FIGURE 5.9: The Hereford Mappa Mundi of 1300 organised places according to their approximate direction and distance from Jerusalem.

Figure 5.10: The SAGE system in use. The SAGE system used light guns as interaction devices.

*Courtesy of Wikipedia. Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.*

FIGURE 5.11: The Whirlwind computer at the MIT Lincoln Laboratory.



### 5.2.1 Summary

Basic diagrammatic conventions rely on quantitative correspondence between a direction on the surface and a continuous quantity such as time or distance. These should follow established conventions of maps and graphs. Where to learn more:

MacEachren, Alan M. (2004): *How Maps Work: Representation, Visualization, and Design*. The Guilford Press

## 5.3 SCHEMATIC DRAWINGS

Ivan Sutherland's groundbreaking PhD research with Whirlwind's successor TX-2 introduced several more sophisticated alternatives (Sutherland 1963). The use of a light pen allowed users to draw arbitrary lines, rather than relying on control keys to select predefined options. An obvious application, in the engineering context of Massachusetts Institute of Technology (MIT) where Sutherland worked, was to make engineering drawings such as the girder bridge in Figure 5.13. Lines on the screen are scaled versions of the actual girders, and text information can be overlaid to give details of force calculations. Plans of this kind, as a visual representation, are closely related to maps. However, where the plane of a map corresponds to a continuous surface, engineering drawings need not be continuous. Each set of connected components must share the same scale, but white space indicates an interpretive break, so that independent representations can potentially share the same divided surface - a convention introduced in Diderot's encyclopedia of 1772, which showed pictures of multiple objects on a page, but cut them loose from any shared pictorial context.

FIGURE 5.12: The TX-2 graphics computer, running Ivan Sutherland's Sketchpad software.

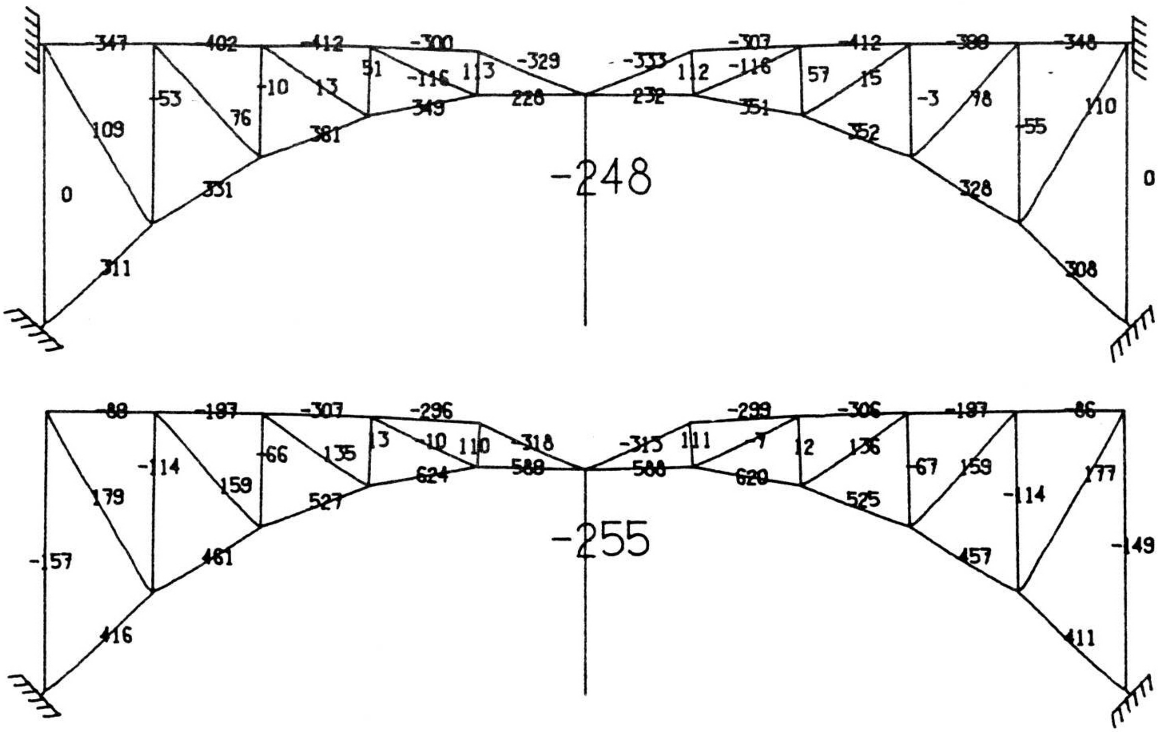



Figure 5.13: An example of a force diagram created using Sutherland's Sketchpad.

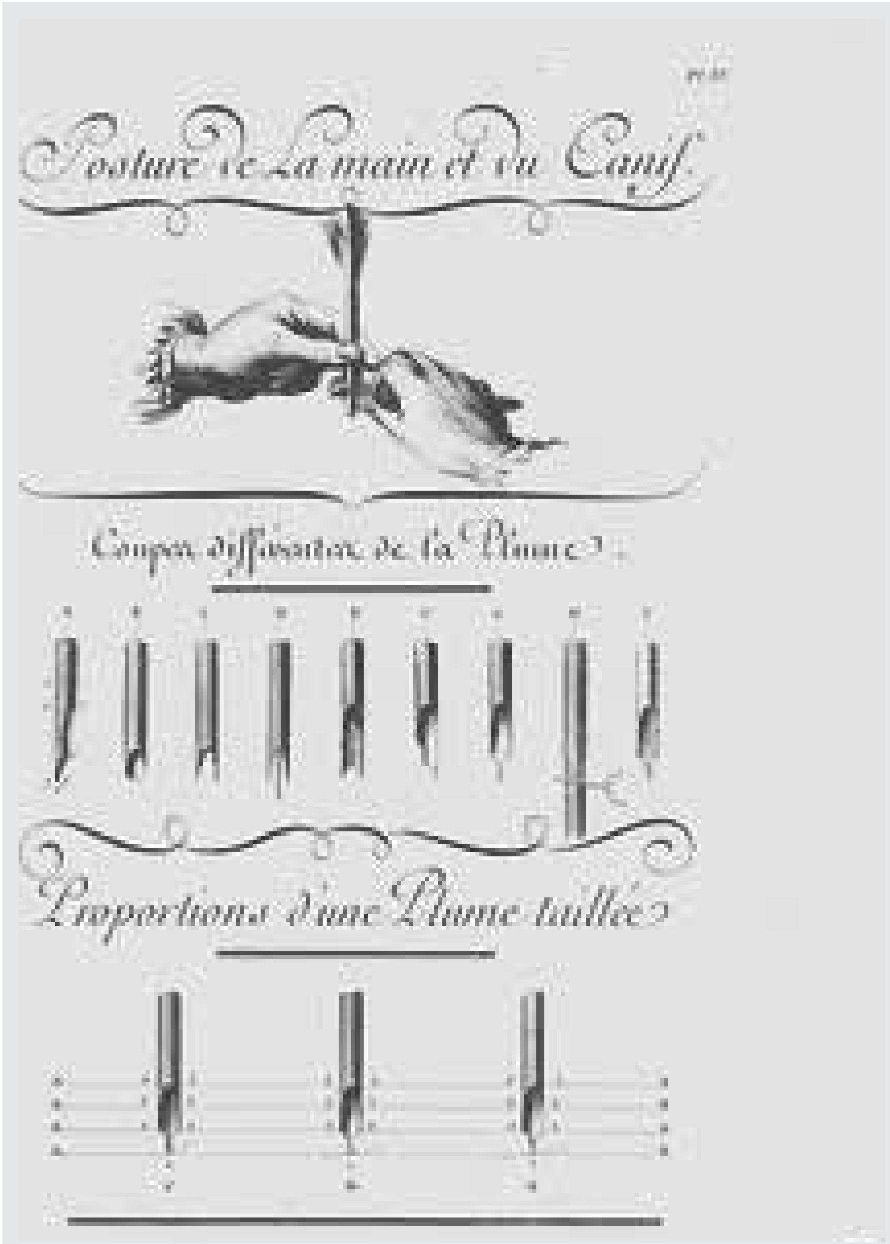



FIGURE 5.14: A page from the Encyclopedie of Diderot and d'Alembert, combining pictorial elements with diagrammatic lines and categorical use of white space.

### 5.3.1 Summary

Engineering drawing conventions allow schematic views of connected components to be shown in relative scale, and with text annotations labelling the parts. White space in the representation plane can be used to help the reader distinguish elements from each other rather than directly representing physical space. Where to learn more:

- Engineering draughting textbooks
- Ferguson, Eugene S. (1994): *Engineering and the Mind's Eye*. MIT Press

## 5.4 PICTURES

The examples so far may seem rather abstract. Isn't the most 'natural' visual representation simply a picture of the thing you are trying to represent? In that case, what is so hard about design? Just point a camera, and take the picture. It seems like pictures are natural and intuitive, and anyone should be able to understand what they mean. Of course, you might want the picture to be more or less artistic, but that isn't a technical concern, is it? Well, Ivan Sutherland also suggested the potential value that computer screens might offer as artistic tools. His Sketchpad system was used to create a simple animated cartoon of a winking girl. We can use this example to ask whether pictures are necessarily 'natural', and what design factors are relevant to the selection or creation of pictures in an interaction design context.

We would not describe Sutherland's girl as 'realistic', but it is an effective representation of a girl. In fact, it is an unusually good representation of a winking girl, because all the other elements of the picture are completely abstract and generic. It uses a conventional graphic vocabulary of lines and shapes that are understood in our culture to represent eyes, mouths and so on - these elements do not draw attention to themselves, and therefore highlight the winking eye. If a realistic picture of an actual person was used instead, other aspects of the image (the particular person) might distract the viewer from this message.

Figure 5.15: Sutherland's 'Winking Girl' drawing, created with the Sketchpad system.



It is important, when considering the design options for pictures, to avoid the 'resemblance fallacy', i.e. that drawings are able to depict real object or scenes because the viewer's perception of the flat image simulates the visual perception of a real scene. In practice, all pictures rely on conventions of visual representation, and are relatively poor simulations of natural engagement with physical objects, scenes and people. We are in the habit of speaking approvingly of some pictures as more 'realistic' than others (photographs, photorealistic ray-traced renderings, 'old master' oil paintings), but this simply means that they follow more rigorously a particular set of conventions. The informed designer is aware of a wide range of pictorial conventions and options.

As an example of different pictorial conventions, consider the ways that scenes can be rendered using different forms of artistic perspective. The invention of linear perspective introduced a particular convention in which the viewer is encouraged to think of the scene as perceived through a lens or frame while holding his head still, so that nearby objects occupy a disproportionate amount of the visual field. Previously, pictorial representations more often varied the relative size of objects according to their importance - a kind of ‘semantic’ perspective. Modern viewers tend to think of the perspective of a camera lens as being most natural, due to the ubiquity of photography, but we still understand and respect alternative perspectives, such as the isometric perspective of the pixel art group eBoy, which has been highly influential on video game style.

FIGURE 5.16: Example of an early work by Masaccio, demonstrating a 'perspective' in which relative size shows symbolic importance.

*Courtesy of Masaccio (1401-1428). Copyright: pd (Public Domain (information that is common property and contains no original authorship)).*


Figure 5.17: Example of the strict isometric perspective used by the eBoy group.

FIGURE 5.18: Masaccio's mature work The Tribute Money, demonstrating linear perspective.



As with most conventions of pictorial representation, new perspective rendering conventions are invented and esteemed for their accuracy by critical consensus, and only more slowly adopted by untrained readers. The consensus on preferred perspective shifts across cultures and historical periods. It would be naïve to assume that the conventions of today are the final and perfect product of technical evolution. As with text, we become so accustomed to interpreting these representations that we are blind to the artifice. But professional artists are fully aware of the conventions they use, even where they might have mechanical elements - the way that a photograph is framed changes its meaning, and a skilled pencil drawing is completely unlike visual edge-detection thresholds. A good pictorial representation need not simulate visual experience any more than a good painting of a unicorn need resemble an actual unicorn. When designing user interfaces, all of these techniques are available for use, and new styles of pictorial rendering are constantly being introduced.

### 5.4.1 Summary

Pictorial representations, including line drawings, paintings, perspective renderings and photographs rely on shared interpretive conventions for their meaning. It is naïve to treat screen representations as though they were simulations of experience in the physical world.

Where to learn more:

- Micklewright, Keith (2005): *Drawing: Mastering the Language of Visual Expression*. Harry N. Abrams
- Stroebel, Leslie, Todd, Hollis and Zakia, Richard (1979): *Visual Concepts for Photographers*. Focal Press

## 5.5 NODE-AND-LINK DIAGRAMS

The first impulse of a computer scientist, when given a pencil, seems to be to draw boxes and connect them with lines. These node and link diagrams can be analysed in terms of the graph structures that are fundamental to the study of algorithms (but unrelated to the visual representations known as graphs or charts). A predecessor of these connectivity diagrams can be found in electrical circuit schematics, where the exact location of components, and the lengths of the wires, can be arranged anywhere, because they are irrelevant to the circuit function. Another early program created for the TX-2, this time by Ivan Sutherland's brother Bert, allowed users to create circuit diagrams of this kind. The distinctive feature of a node-and-link connectivity diagram is that, since the position of each node is irrelevant to the operation of the circuit, it can be used to carry other information. Marian Petre's research into the work of electronics engineers (Petre 1995) catalogued the ways in which they positioned components in ways that were meaningful to human readers, but not to the computer - like the blank space between Diderot's objects this is a form of 'secondary notation' - use of the plane to assist the reader in ways not related to the technical content.

Circuit connectivity diagrams have been most widely popularised through the London Underground diagram, an invention of electrical engineer Henry Beck. The diagram clarified earlier maps by exploiting the fact that most underground travellers are only interested in order and connectivity, not location, of the stations on the line. (Sadly, the widespread belief that a 'diagram' will be technical and hard to understand means that most people describe this as the London Undergound 'map', despite Beck's insistence on his original term).

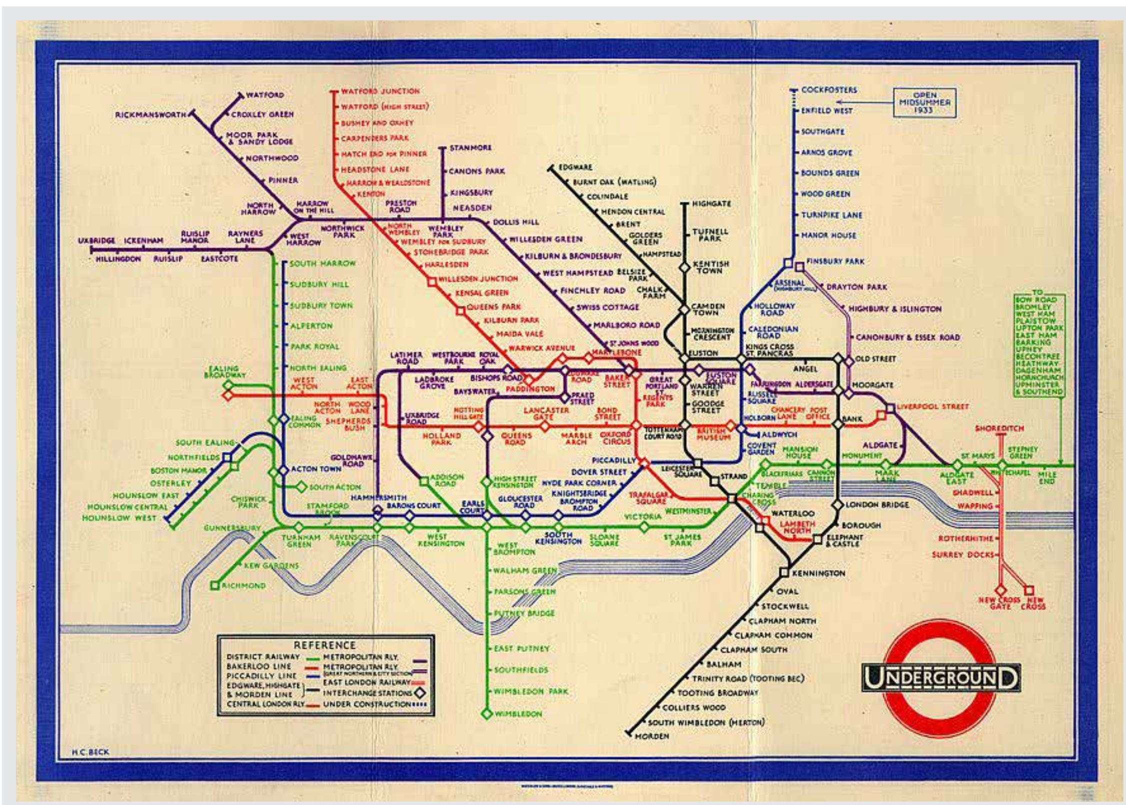



FIGURE 5.19: Henry Beck's London Underground Diagram (1933).

*Courtesy of Harry C. Beck and possibly F. H. Stingemore, born 1890, died 1954. Stingmore designed posters for the Underground Group and London Transport 1914-1942. Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.*

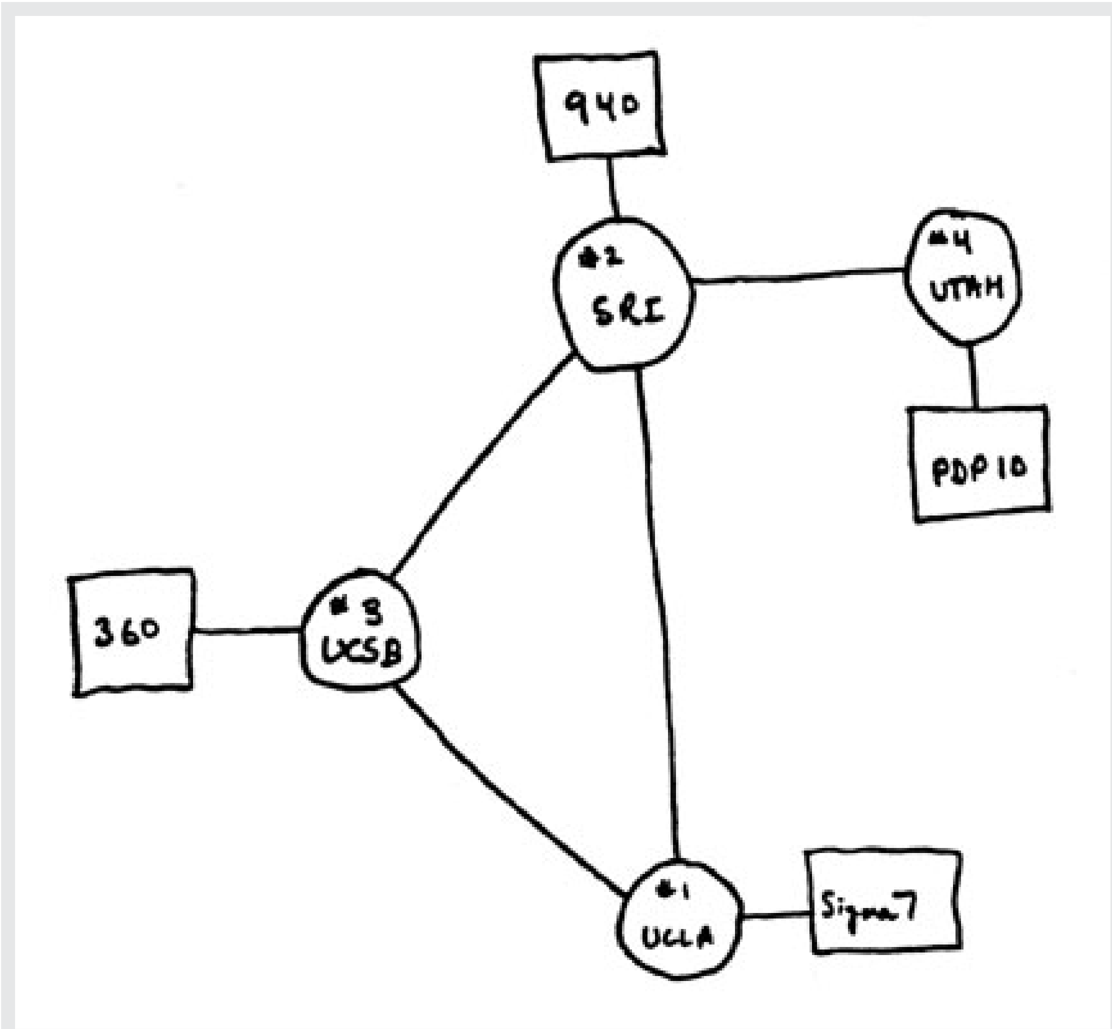



FIGURE 5.20: Node and link diagram of the kind often drawn by computing professionals.

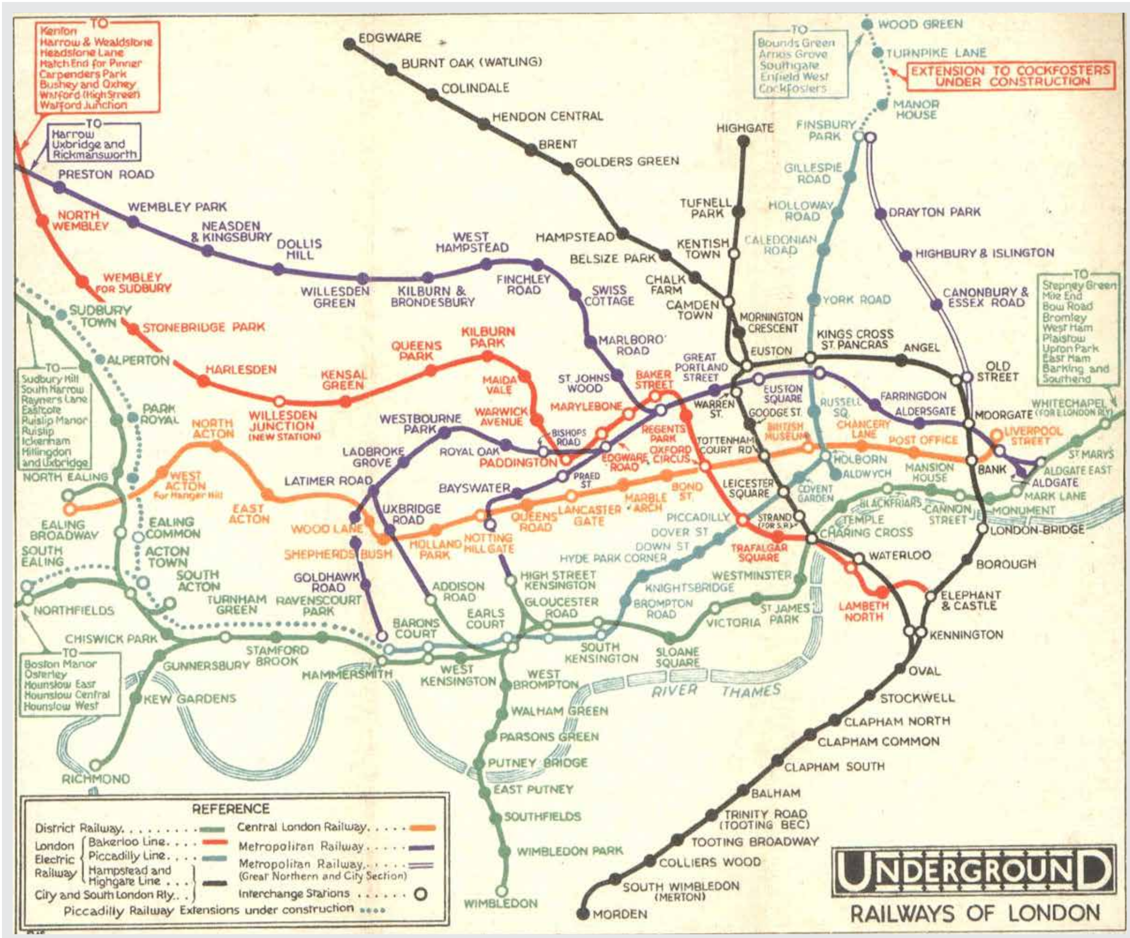



Figure 5.21: Map of the London Underground network, as it was printed before the design of Beck's diagram (1932).

*Courtesy of Harry C. Beck and possibly F. H. Stingemore, born 1890, died 1954. Stingmore designed posters for the Underground Group and London Transport 1914-1942. Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.*

## 5.5.1 Summary

Node and link diagrams are still widely perceived as being too technical for broad acceptance. Nevertheless, they can present information about ordering and relationships clearly, especially if consideration is given to the value of allowing human users to specify positions.

Where to learn more:

- Diagrammatic representation books
- Lowe, Ric (1992): *Successful Instructional Diagram.*

## 5.6 ICONS AND SYMBOLS

Maps frequently use symbols to indicate specific kinds of landmark. Sometimes these are recognisably pictorial (the standard symbols for tree and church), but others are fairly arbitrary conventions (the symbol for a railway station). As the resolution of computer displays increased in the 1970s, a greater variety of symbols could be differentiated, by making them more detailed, as in the MIT SDMS (Spatial Data Management System) that mapped a naval battle scenario with symbols for different kinds of ship. However, the dividing line between pictures and symbols is ambiguous. Children's drawings of houses often use conventional symbols (door, four windows, triangle roof and chimney) whether or not their own house has two storeys, or a fireplace. Letters of the Latin alphabet are shapes with completely arbitrary relationship to their phonetic meaning, but the Korean phonetic alphabet is easier to learn because the forms mimic the shape of the mouth when pronouncing those sounds. The field of semiotics offers sophisticated ways of analysing the basis on which marks correspond to meanings. In most cases, the best approach for an interaction designer is simply to adopt familiar conventions. When these do not exist, the design task is more challenging.

It is unclear which of the designers working on the Xerox Star coined the term 'icon' for the small pictures symbolising different kinds of system object. David Canfield Smith winningly described them as being like religious icons, which he said were pictures standing for (abstract) spiritual concepts. But 'icon' is also used as a technical term in semiotics. Unfortunately, few of the Xerox team had a sophisticated understanding of semiotics. It was fine art PhD Susan Kare's design work on the Apple Macintosh that established a visual vocabulary which has informed the genre ever since. Some general advice principles are offered by authors such as Horton (1994), but the successful design of icons is still sporadic. Many software publishers simply opt for a memorable brand logo, while others seriously misjudge the kinds of correspondence that are appropriate (my favourite blooper was a software engineering tool in which a pile of coins was used to access the 'change' command).

It has been suggested that icons, being pictorial, are easier to understand than text, and that pre-literate children, or speakers of different languages, might thereby be able to use computers without being able to read. In practice, most icons simply add decoration to text labels, and those that are intended to be self-explanatory must be supported with textual tooltips. The early Macintosh icons, despite their elegance, were surprisingly open to misinterpretation. One PhD graduate of my acquaintance believed that the Macintosh folder symbol was a briefcase (the folder tag looked like a handle), which allowed her to carry her files from place to place when placed inside it. Although mistaken, this belief never caused her any trouble - any correspondence can work, so long as it is applied consistently.

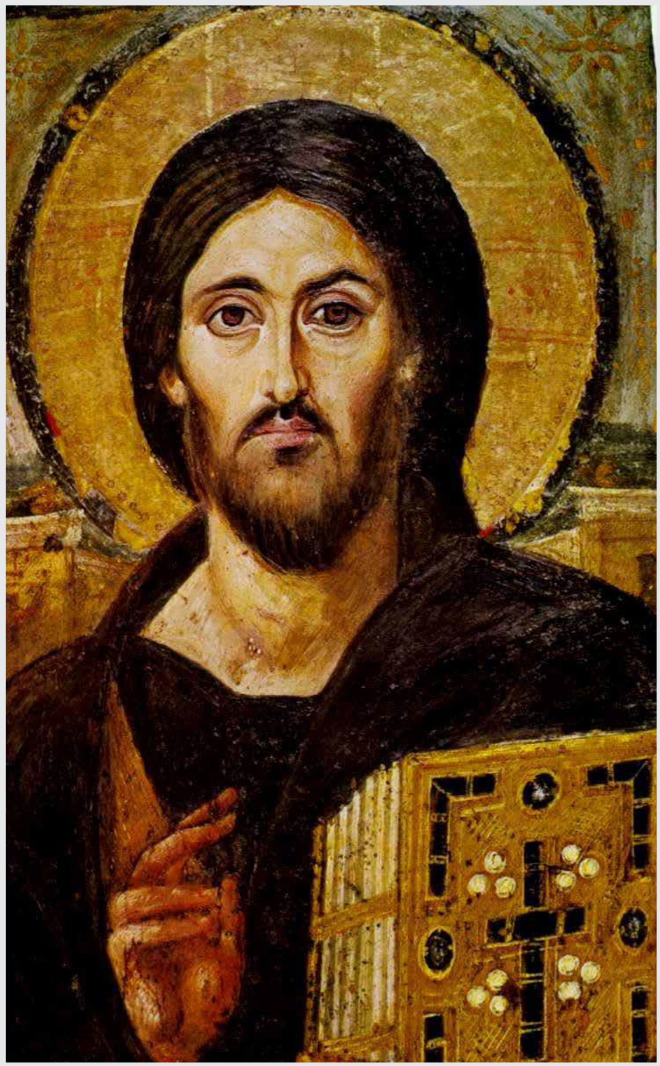


FIGURE 5.22: In art, the term Icon (from Greek, eikon, “image”) commonly refers to religious paintings in Eastern Orthodox, Oriental Orthodox, and Eastern-rite Catholic jurisdictions. Here a 6th-century encaustic icon from Saint Catherine’s Monastery, Mount Sinai.

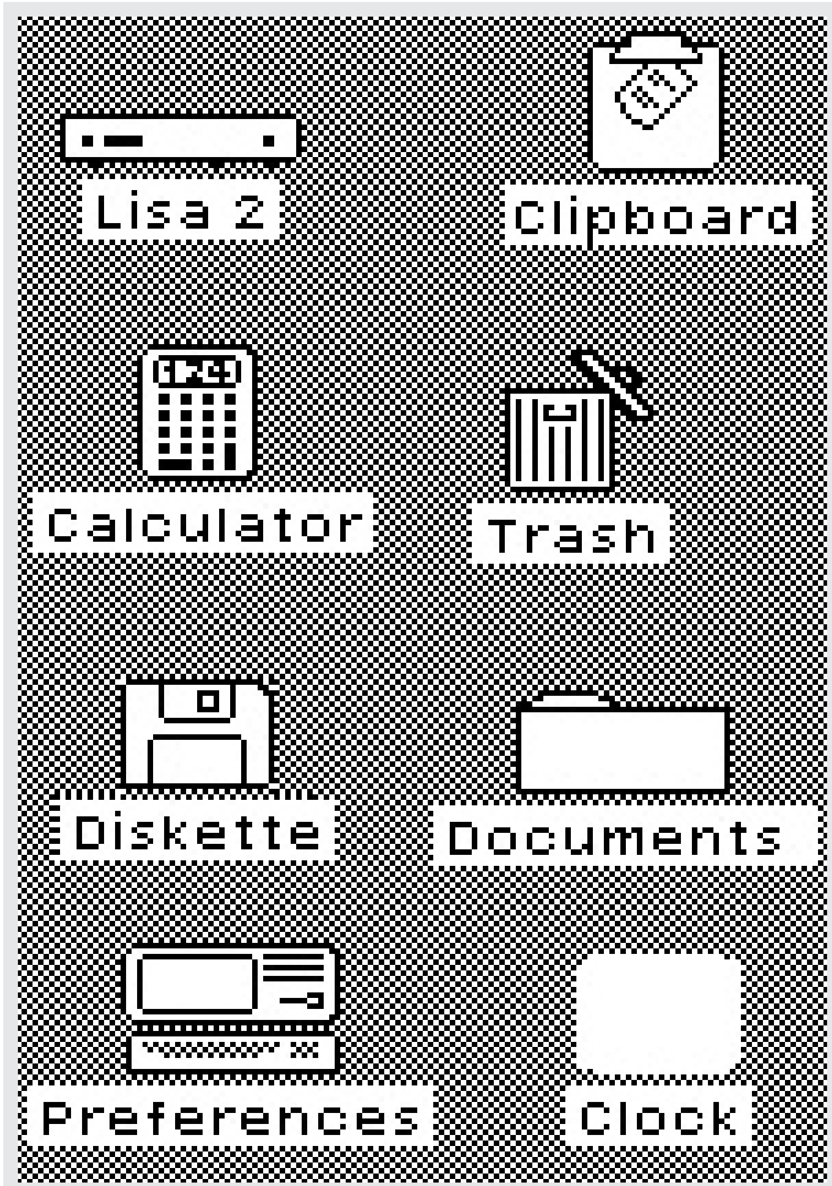



FIGURE 5.23: In computing, David Canfield Smith described computer icons as being like religious icons, which he said were pictures standing for (abstract) spiritual concepts.

### 5.6.1 Summary

The design of simple and memorable visual symbols is a sophisticated graphic design skill. Following established conventions is the easiest option, but new symbols must be designed with an awareness of what sort of correspondence is intended - pictorial, symbolic, metonymic (e.g. a key to represent locking), bizarrely mnemonic, but probably not monolingual puns.
Where to learn more:

Napoles, Veronica (1987): *Corporate Identity Design*.

## 5.7 VISUAL METAPHOR

The ambitious graphic designs of the Xerox Star/Alto and Apple Lisa/Macintosh were the first mass-market visual interfaces. They were marketed to office professionals, making the ‘cover story’ that they resembled an office desktop a convenient explanatory device. Of course, as was frequently noted at the time, these interfaces behaved nothing like a real desktop. The mnemonic symbol for file deletion (a wastebasket) was ridiculous if interpreted as an object placed on a desk. And nobody could explain why the desk had windows in it (the name was derived from the ‘clipping window’ of the graphics architecture used to implement them - it was at some later point that they began to be explained as resembling sheets of paper on a desk). There were immediate complaints from luminaries such as Alan Kay and Ted Nelson that strict analogical correspondence to physical objects would become obstructive rather than instructive. Nevertheless, for many years the marketing story behind the desktop metaphor was taken seriously, despite the fact that all attempts to improve the Macintosh design with more elaborate visual analogies, as in General Magic and Microsoft Bob, subsequently failed.

The ‘desktop’ can be far more profitably analysed (and extended) by understanding the representational conventions that it uses. The size and position of icons and windows on the desktop has no meaning, they are not connected, and there is no visual perspective, so it is neither a map, graph nor picture. The real value is the extent to which it allows secondary notation, with the user creating her own meaning by arranging items as she wishes. Window borders separate areas of the screen into different pictorial, text or symbolic contexts as in the typographic page design of a textbook or magazine. Icons use a large variety of conventions to indicate symbolic correspondence to software operations and/or company brands, but they are only occasionally or incidentally organised into more complex semiotic structures.

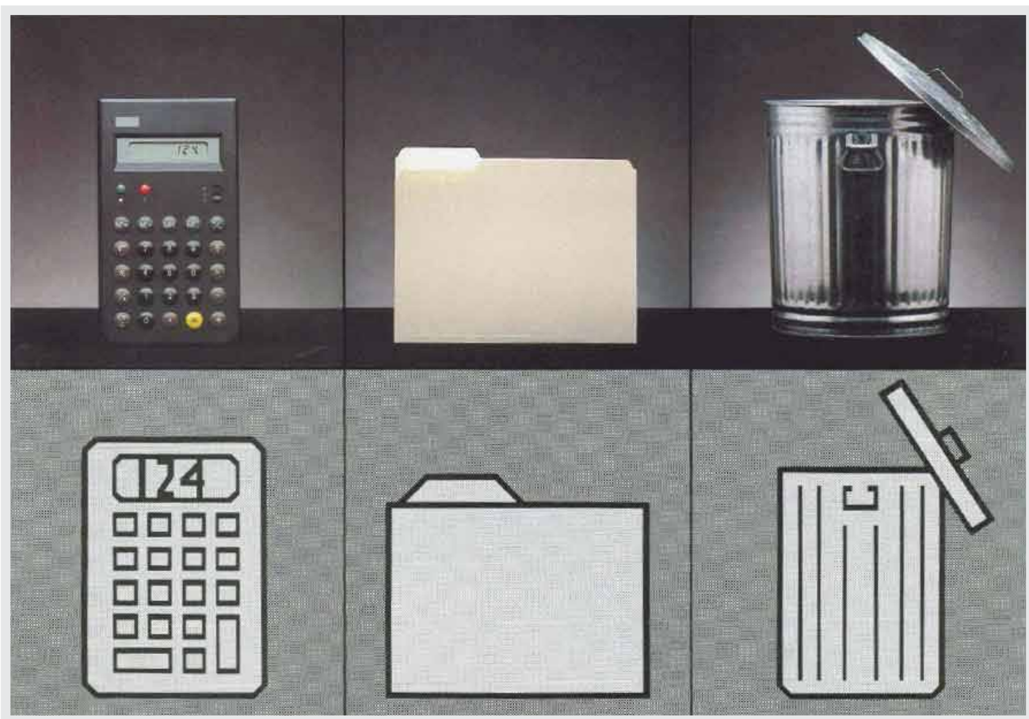



FIGURE 5.24: Apple marketed the visual metaphor in 1983 as a key benefit of the Lisa computer. This advertisement said 'You can work with Lisa the same familiar way you work at your desk'. However a controlled study by Carroll and Mazur (1986) found that the claim for immediately familiar operation may have been exaggerated.

FIGURE 5.25: The Xerox Alto and Apple Lisa, early products in which bitmapped displays allowed pictorial icons to be used as mnemonic cues within the ‘desktop metaphor’

FIGURE 5.26: Apple Lisa.

*Courtesy of Mschlindwein. Copyright: CC-Att-SA (Creative Commons Attribution-ShareAlike 3.0 Unported).*

## 5.7.1 Summary

Theories of visual representation, rather than theories of visual metaphor, are the best approach to explaining the conventional Macintosh/Windows 'desktop'. There is huge room for improvement.

Where to learn more:

Blackwell, Alan (2006): *The reification of metaphor as a design tool.* In ACM Transactions on Computer-Human Interaction, 13 (4) pp. 490-530

## 5.8 UNIFIED THEORIES OF VISUAL REPRESENTATION

The analysis in this article has addressed the most important principles of visual representation for screen design, introduced with examples from the early history of graphical user interfaces. In most cases, these principles have been developed and elaborated within whole fields of study and professional skill - typography, cartography, engineering and architectural draughting, art criticism and semiotics. Improving on the current conventions requires serious skill and understanding. Nevertheless, interaction designers should be able, when necessary, to invent new visual representations.

One approach is to take a holistic perspective on visual language, information design, notations, or diagrams. Specialist research communities in these fields address many relevant factors from low-level visual perception to critique of visual culture. Across all of them, it can be necessary to ignore (or not be distracted by) technical and marketing claims, and to remember that all visual representations simply comprise marks on a surface that are intended to correspond to things understood by the reader. The two dimensions of the surface can be made to correspond to physical space (in a map), to dimensions of an object, to a pictorial perspective, or to continuous abstract scales (time or quantity). The surface can also be partitioned into regions that should be interpreted differently. Within any region, elements can be aligned, grouped, connected or contained in order to express their relationships. In each case, the correspondence between that arrangement, and the intended interpretation, must be understood by convention, explained, or derived from the structural and perceptual properties of marks on the plane. Finally, any individual element might be assigned meaning according to many different semiotic principles of correspondence.

The following table summarises holistic views, as introduced above, drawing principally on the work of Bertin, Richards, MacEachren, Blackwell & Engelhardt and Engelhardt.

Where to learn more:

Engelhardt, Yuri (2002). *The Language of Graphics. A framework for the analysis of syntax and meaning in maps, charts and diagrams (PhD Thesis).* University of Amsterdam

| | **Graphic Resources** | **Correspondence** | **Design Uses** |
|---|---|---|---|
| Marks | Shape<br>Orientation<br>Size<br>Texture<br>Saturation<br>Colour<br>Line | Literal<br>(visual imitation of physical features)<br>Mapping<br>(quantity, relative scale)<br>Conventional<br>(arbitrary) | Mark position, identify category (shape, texture colour)<br>Indicate direction (orientation, line)<br>Express magnitude (saturation, size, length)<br>Simple symbols and colour codes |
| Symbols | Geometric elements<br>Letter forms<br>Logos and icons<br>Picture elements<br>Connective elements | Topological (linking)<br>Depictive<br>(pictorial conventions)<br>Figurative (metonym, visual puns) Connotative (professional and cultural association) Acquired (specialist literacies) | Texts and symbolic calculi<br>Diagram elements<br>Branding<br>Visual rhetoric<br>Definition of regions |

| | | | |
|---|---|---|---|
| Regions | Alignment grids<br>Borders and frames<br>Area fills<br>White space<br>Gestalt integration | Containment<br>Separation<br>Framing (composition, photography) Layering | Identifying shared membership<br>Segregating or nesting multiple surface conventions in panels Accommodating labels, captions or legends |
| Surfaces | The plane<br>Material object on which the marks are imposed (paper, stone) Mounting, orientation and display context Display medium | Literal (map)<br>Euclidean<br>(scale and angle) Metrical (quantitative axes)<br>Juxtaposed or ordered (regions, catalogues)<br>Image-schematic<br>Embodied/situated | Typographic layouts<br>Graphs and charts<br>Relational diagrams<br>Visual interfaces<br>Secondary notations<br>Signs and displays |

Table 5.1: Summary of the ways in which graphical representations can be applied in design, via different systems of correspondence.

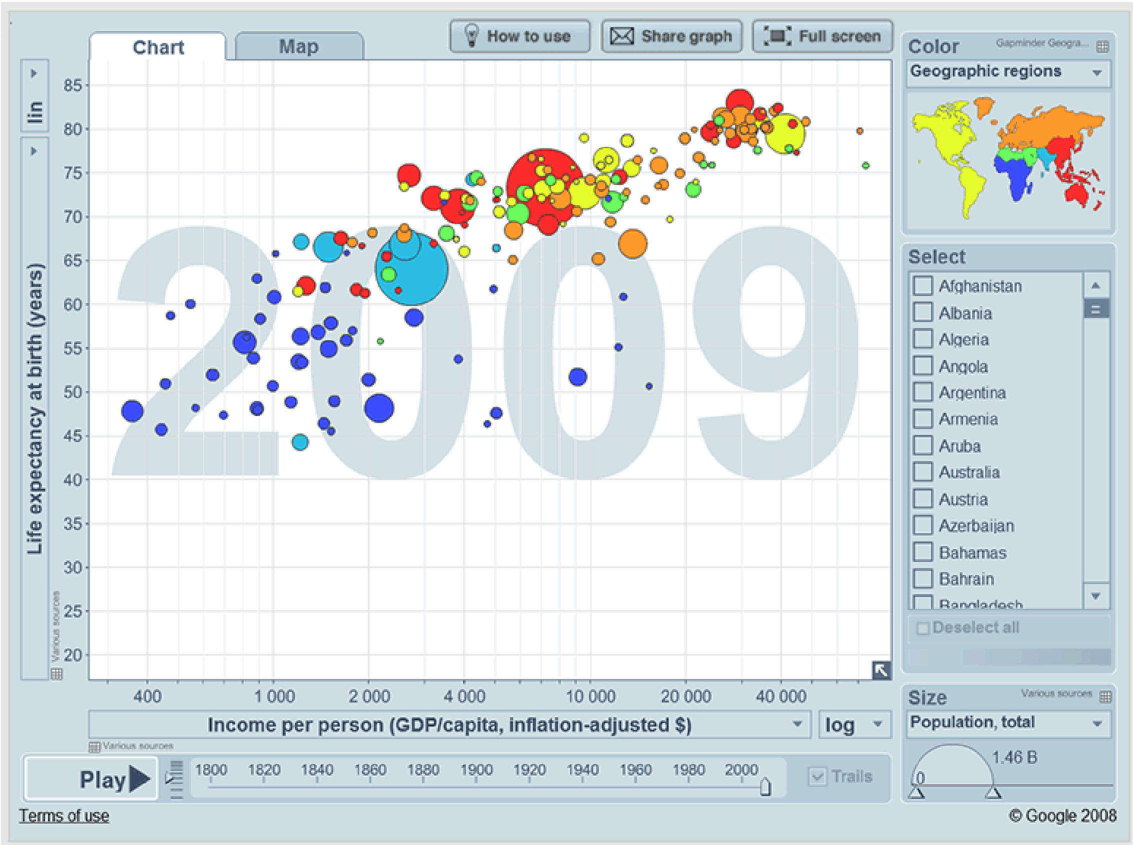



Table 5.2: Screenshot from the site gapminder.org, illustrating a variety of correspondence conventions used in different parts of the page.

As an example of how one might analyse (or working backwards, design) a complex visual representation, consider the case of musical scores. These consist of marks on a paper surface, bound into a multi-page book, that is placed on a stand at arms length in front of a performer. Each page is vertically divided into a number of regions, visually separated by white space and grid alignment cues. The regions are ordered, with that at the top of the page coming first. Each region contains two quantitative axes, with the horizontal axis representing time duration, and the vertical axis pitch. The vertical axis is segmented by lines to categorise pitch class. Symbols placed at a given x-y location indicate a specific pitched sound to be initiated at a specific time. A conventional symbol set indicates the duration of the sound. None of the elements use any variation in colour, saturation or texture. A wide variety of text labels and annotation symbols are used to elaborate these basic elements. Music can be, and is, also expressed using many other visual representations (see e.g. Duignan for a survey of representations used in digital music processing).

## 5.9 WHERE TO LEARN MORE

The historical examples of early computer representations used in this article are mainly drawn from Sutherland (Ed. Blackwell and Rodden 2003), Garland (1994), and Blackwell (2006). Historical reviews of visual representation in other fields include Ferguson (1992), Pérez-Gómez and Pelletier (1997), McCloud (1993), Tufte (1983). Reviews of human perceptual principles can be found in Gregory (1970), Ittelson (1996), Ware (2004), Blackwell (2002). Advice on principles of interaction with visual representation is distributed throughout the HCI literature, but classics include Norman (1988), Horton (1994), Shneiderman ( Shneiderman and Plaisant 2009, Card et al 1999, Bederson and Shneiderman 2003) and Spence (2001). Green's Cognitive Dimensions of Notations framework has for many years provided a systematic classification of the design parameters in interactive visual representations. A brief introduction is provided in Blackwell and Green (2003).

Research on visual representation topics is regularly presented at the Diagrams conference series (which has a particular emphasis on cognitive science), the InfoDesign and Vision Plus conferences (which emphasise graphic and typographic information design), the Visual Languages and Human-Centric Computing symposia (emphasising software tools and development), and the InfoVis and Information Visualisation conferences (emphasising quantitative and scientific data visualisation).

### IV - International Conference on Information Visualization

2008 2007 2006 2005 2004 2003 2002 2001
2000 1999 1998

### DIAGRAMS - International Conference on the Theory and Application of Diagrams

2008 2006 2004 2002 2000

### VL-HCC - Symposium on Visual Languages and Human Centric Computing

2008 2007 2007 2006 2005 2004 2003 2002
2001 2000 1999 1998 1997 1996 1995 1994
1993 1992 1991 1990

Next conference is coming up 15 Sep 2013 in San Jose, CA, USA

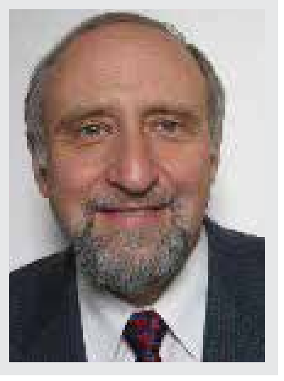
### InfoVis - IEEE Symposium on Information Visualization

2005 2004 2003 2002 2001 2000 1999 1998
1997 1995

## 5.11 COMMENTARY BY BEN SHNEIDERMAN

How to cite this commentary in your report

### Ben Shneiderman



Ben Shneiderman is a Professor in the Department of Computer Science, Founding Director (1983-2000) of the Human-Computer Interaction Laboratory, and Member of the Institute for Advanced Computer Studies at the University of Maryland at College Park. He has taught previously at the State University of New York and at Indiana University. He was made a Fellow of the ACM in 1997, elected a ...

Ben Shneiderman
Ben Shneiderman is a member of The Interaction Design Foundation

Since computer displays are such powerful visual appliances, careful designers devote extensive effort to getting the visual representation right. They have to balance the demands of many tasks, diverse users, and challenging requirements, such as short learning time, rapid performance, low error rates, and good retention over time. Designing esthetic interfaces that please and even delight users is a further expectation that designers must meet to be successful. For playful and discretionary tasks esthetic concerns may dominate, but for life critical tasks, rapid performance with low error rates are essential.

Alan Blackwell's competent description of many visual representation issues is a great start for newcomers with helpful reminders even for experienced designers. The videos make for a pleasant personal accompaniment that bridges visual representation for interface design with thoughtful analyses of representational art. Blackwell's approach might be enriched by more discussion of visual representations in functional product design tied to meaningful tasks. Learning from paintings of Paris is fine, but aren't there other lessons to learn from visual representations in airport kiosks, automobile dashboards, or intensive care units?

These devices as well as most graphical user interfaces and mobile devices raise additional questions of changing state visualization and interaction dynamics. Modern designers need to do more than show the right phone icon, they need to show ringing, busy, inactive, no network, conference mode, etc., which may include color changes (highlighted, grayed out), animations, and accompanying sounds. These designers also need to deal with interactive visual representations that happen with a click, double-click, right-click, drag, drag-and-drop, hover, multi-select, region-select, brushing-linking, and more.

The world of mobile devices such as phones, cameras, music players, or medical sensors is the new frontier for design, where visual representations are dynamic and tightly integrated with sound, haptics, and novel actions such as shaking, twisting, or body movements. Even more challenging is the expectation that goes beyond the solitary viewer to the collaboration in which multiple users embedded in a changing physical environment produce new visual representations.

These changing and interactive demands on designers invite creative expressions that are very different from designs for static signs, printed diagrams, or interpretive art. The adventure for visual representation designers is to create a new language of interaction that engages users, accelerates learning, provides comprehensible feedback, and offers appropriate warnings when dangers emerge. Blackwell touches on some of these issues in the closing Gapminder example, but I was thirsty for more.

## 5.12 COMMENTARY BY CLIVE RICHARDS

How to cite this commentary in your report

### Clive Richards

© *Clive Richards*

Clive Richards, MPhil PhD(RCA) FCSD FRSA, is a Visiting Professor to the Faculty of Arts and Architecture, University of Brighton. He has worked in a wide range of art and design fields, including commercial practice (technical illustration, information graphics, corporate identity, typography and artists' catalogue design), academic research, undergraduate and postgraduate teaching, PhD...

Clive Richards
Clive Richards is a member of The Interaction Design Foundation

If I may be permitted a graphically inspired metaphor Alan Blackwell provides us with a neat pen sketch of that extensive scene called 'visual representation' (Blackwell 2011).

"Visualisation has a lot more to offer than most people are aware of today" we are told by Robert Kosara at the end of his commentary (Kosara 2010) on

Stephen Few's related article on 'Data visualisation for human perception' (Few 2010). Korsara is right, and Blackwell maps out the broad territory in which many of these visualisation offerings may be located. In this commentary I offer a few observations on some prominent features in that landscape: dynamics, picturing, semiotics and metaphor.

## Dynamics

Ben Shneiderman's critique of Blackwell's piece points to a lack of attention to "... additional questions of changing state visualisations and interaction dynamics" (Shneiderman 2010). Indeed the possibilities offered by these additional questions present some exciting challenges for interaction designers - opportunities to create novel and effective combinations of visual with other sensory and motor experiences in dynamic operational contexts. Shneiderman suggests that: "These changing and interactive demands on designers invite creative expressions that are very different from design for static signs, printed diagrams, or interpretive art". This may be so up to a point, but here Shneinderman and I part company a little. The focus of Blackwell's essay is properly on the visual representation side of facilities available to interaction designers, and in that context he is quite right to give prominence to highly successful but static visual representation precedents, and also to point out the various specialist fields of endeavour in which they have been developed. Some of these representational approaches have histories reaching back thousands of years and are deeply embedded within our culture. It would be foolhardy to disregard conventions established in, say, the print domain, and to try to re-invent everything afresh for the screen, even if this were a practical proposition. Others have made arguments to support looking to historical precedents. For example Michael Twyman has pointed out that when considering typographic cueing and "... the problems of the electronic age ... we have much to learn from the manuscript age" (Twyman 1987, p5). He proposes that studying the early scribes' use of colour, spacing and other graphical devices can usefully inform the design of today's screen-based texts. And as Blackwell points out in his opening section on 'Typography and text' "most information on computer screen is still presented as text".

It is also sometimes assumed that the pictorial representation of a dynamic process is best presented dynamically. However it can be argued that the comic book convention of using a sequence of static frames is sometimes superior for focusing the viewer's attention on the critical events in a process, rather than using an animated sequence in which key moments may be missed. This is of course not to deny the immense value of the moving and interactive visual image in the right context. The Gapminder charts are a case in point (http://www.gapminder.org). Blackwell usefully includes one of these, but as a static presentation. These diagrams come to life and really tell their story through the clustering of balloons that inflate or deflate as they move about the screen when driven through simulated periods of time.

While designing a tool for engineers to learn about the operation and maintenance of an oil system for an aircraft jet engine, Detlev Fischer devised a series of interactive animations, called 'Cinegrams' to display in diagrammatic form various operating procedures (Fischer and Richards 1995). He used the cinematic techniques of time compression and expansion in one animated sequence to show how the slow accumulation of debris in an oil filter, over an extended period of time, would eventually create a blockage to the oil flow and trigger the opening of a by-pass device in split seconds. Notwithstanding my earlier comment about the potential superiority of the comic strip genre for displaying some time dependant processes this particular Cinegram proved very instructive for the targeted users. There are many other examples one could cite where dynamic picturing of this sort has been deployed to similarly good effect in interactive environments.

## Picturing

Shneinderman also comments that: "Blackwell's approach might be enriched by more discussion of visual representation in functional product design tied to meaningful tasks". An area I have worked in is the pictorial representation of engineering assemblies to show that which is normally hidden from view. Techniques to do this on the printed page include 'ghosting' (making occluding parts appear as if transparent), 'exploding' (showing components separately, set out in dis-assembly order along an axis) and cutting away (taking a slice out of an outer shell to reveal mechanisms beneath). All these three-dimensional picturing techniques were used by, if not actually invented by, Leonardo Da Vinci (Richards 2006). All could be enhanced by interactive viewer control - an area of further fruitful exploration for picturing purposes in technical documentation contexts.

Blackwell's section on 'Pictures' warns us that when considering picturing options to avoid the "resemblance fallacy" pointing out the role that convention plays, even in so called photo-realistic images. He also points out that viewers can be distracted from the message by incidental information in 'realistic' pictures. From my own work in the field I know that technical illustrators' synoptic black and white outline depictions are regarded as best for drawing the viewer's attention to the key features of a pictorial representation. Research in this area has shown that when using linear perspective type drawings the appropriate deployment of lines of varying 'weight', rather than of a single thickness, can have a significant effect on viewers' levels of understanding about what is depicted (Richards, Bussard and Newman 2007). This work was done specifically to determine an 'easy to read' visual representational style when manipulating on the screen images of CAD objects. The most effective convention was shown to be: thin lines for edges where both planes forming the edge are visible and thicker lines for edges where only one plane is visible - that is where an outline edge forms a kind of horizon to the object.

These line thickness conventions appear on the face of it to have little to do with how we normally perceive the world, and Blackwell tells us that: "A good pictorial representation need not simulate visual experience any more than a good painting of a unicorn need resemble an actual unicorn". And some particular representations of unicorns can aid our understanding of how to use semiotic theory to figure out how pictures may be interpreted and, importantly, sometimes misunderstood - as I shall describe in the following.

## Semiotics

Blackwell mentions semiotics, almost in passing, however it can help unravel some of the complexities of visual representation. Evelyn Goldsmith uses a Charles Addams cartoon to explain the relevance of the 'syntactic', 'semantic' and 'pragmatic' levels of semiotic analysis when applied to pictures (Goldsmith 1978). The cartoon in question, like many of those by Charles Addams, has no caption. It shows two unicorns standing on a small island in the pouring rain forlornly watching the Ark sailing away into the distance. Goldsmith suggests that most viewers will have little trouble in interpreting the overlapping elements in the scene, for example that one unicorn is standing behind the other, nor any difficulty understanding that the texture gradient of the sea stands for a receding horizontal plane. These represent the syntactic level of interpretation. Most adults will correctly identify the various components of the picture at the semantic level, however Goldsmith proposes that a young child might mistake the unicorns for horses and be happy with 'boat' for the Ark. But at the pragmatic level of interpretation, unless a viewer of the picture is aware of the story of Noah's Ark, the joke will be lost - the connection will not be made between the scene depicted in the drawing and the scarcity of unicorns. This reinforces the point that one should not assume that the understanding of pictures is straightforward. There is much more to it than a simple matter or recognition. This is especially the case when metaphor is involved in visual representation.

## Metaphor

Blackwell's section on 'Visual metaphor' is essentially a critique of the use of "theories of visual metaphor" as an "approach to explaining the conventional Mackintosh/Windows 'desktop' ". His is a convincing argument but there is much more which may be said about the use of visual metaphor - especially to show that which otherwise cannot be pictured. In fact most diagrams employ a kind of spatial metaphor when not depicting physical arrangements, for example when using the branches of a tree to represent relations within a family (Richards 2002). The capability to represent the invisible is the great strength of the visual metaphor, but there are dangers, and here I refer back to semiotics and particularly the pragmatic level of analysis. One needs to know the story to get the picture.

In our parental home, one of the many books much loved by my two brothers and me, was *The Practical Encyclopaedia for Children* (Odhams circa 1948). In it a double page spread illustration shows the possible evolutionary phases of the elephant. These are depicted as a procession of animals in a primordial swamp cum jungle setting. Starting with a tiny fish and passing to a small aquatic creature climbing out of the water onto the bank the procession progresses on through eight phases of transformation, including the Moeritherium and the Paleomatodon, finishing up with the land-based giant of today's African Elephant. Recently one of my brothers confessed to me that through studying this graphical diorama he had believed as a child that the elephant had a life cycle akin to that of a frog. He had understood that the procession was a metaphor for time. He had just got the duration wrong - by several orders of magnitude. He also hadn't understood that each separate depiction was of a different animal. He had used the arguably more sophisticated concept that it was the same animal at different times and stages in its individual development.

Please forgive the cliché if I say that this anecdote clearly illustrates that there can be more to looking at a picture than meets the eye? Blackwell's essay provides some useful pointers for exploring the possibilities of this fascinating territory of picturing and visual representation in general.

## References

- Blackwell A 2011 ‘Visual representation’ *Interaction-Design.org*
- Few S 2010 ‘Data visualisation for human perception’ *Interaction-Design.org*
- Fischer D and Richards CJ 1995 ‘The presentation of time in interactive animated systems diagrams’ In: Earnshaw RA and Vince JA (eds) *Multimedia Systems and Applications* London: Academic Press Ltd (pp141 - 159). ISBN 0-12-227740-6
- Goldsmith E 1978 *An analysis of the elements affecting comprehensibility of illustrations intended as supportive of text*PhD thesis (CNAA) Brighton Polytechnic
- Korsa R 2010 ‘Commentary on Stephen Few’s article: Data visualisation for human perception’ *Interaction-Design.org*
  Odhams c. 1949 *The practical encyclopaedia for children*
  (pp 194 - 195)
- Richards CJ 2002 ‘The fundamental design variables of diagramming’ In: Oliver P, Anderson M and Meyer B (eds) *Diagrammatic representation and reasoning* London: Springer Verlag (pp 85 - 102) ISBN 1-85233-242-5
- Richards CJ 2006 ‘Drawing out information - lines of communication in technical illustration’ *Information Design Journal* 14 (2) 93 - 107
- Richards CJ, Bussard N, Newman R 2007 ‘Weighing up line weights: the value of differing line thicknesses in technical illustrations’ *Information Design Journal* 15 (2) 171 - 181
- Shneiderman B 2011 ‘Commentary on Alan Blackwell’s article:

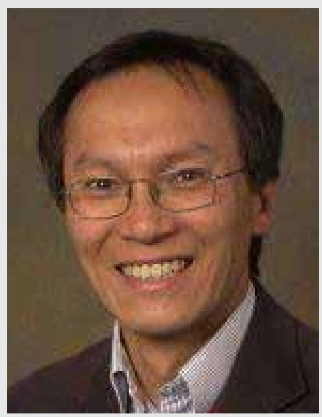
Visual representation' *Interaction-Design.org*

- Twyman M 1982 'The graphic representation of language' *Information Design Journal* 3 (1) 2 - 22

## 5.13 COMMENTARY BY PETER C-H. CHENG

How to cite this commentary in your report

### Peter C-H. Cheng



© Peter C-H. Cheng

Peter Cheng's main research interest is in the nature of representational systems, spanning the design of external representations and the cognitive processes that deal with internal mental representations. Knowledge rich representations used for higher forms of cognition (e.g., complex problem solving, discovery, conceptual learning) in conceptually demanding domains are a particular i...

Peter C-H. Cheng
Peter C-H. Cheng is a member of The Interaction Design Foundation

Alan Blackwell has provided us with a fine introduction to the design of visual representations. The article does a great job in motivating the novice designer of visual representations to explore some of the fundamental issues that lurk just beneath the surface of creating effective representations. Furthermore, he gives us all quite a challenge:

> "*Improving on the current conventions requires serious skill and understanding. Nevertheless, interaction designers should be able, when necessary, to invent new visual representations.*"

Alan, quite rightly, claims that we must consider the fundamental principles of symbolic correspondence, if we are to design new genres of visual representations beyond the common forms of displays and interfaces. The report begins to equip the novice visual representation designer with an understanding of the nature of symbolic correspondence between the components of visual representations and the things they represent, whether objects, actions or ideas. In particular, it gives a useful survey of how correspondence works in a range of representations and provides a systematic framework of how systems of correspondence can be applied to design. The interactive screen shot is an exemplary visual representation that vividly reveals the correspondence techniques used in each part of the example diagram.

However, suppose you really wished to rise to the challenge of creating novel visual representations, how far will a knowledge of the fundamentals of symbolic correspondence take you? Drawing on my studies of the role of diagrams in the history of science, experience of inventing novel visual representations and research on problem solving and learning with diagrams, from the perspective of Cognitive Science, my view is that such knowledge will be necessary but not sufficient for your endeavours. So, what else should the budding visual representation designer consider? From the perspective of cognitive science there are at least three aspects that we may profitably target.

First, there is the knowledge of how human process information; specifically the nature of the human cognitive architecture. By this, I mean more than visual perception, but an understanding of how we mentally receive, store, retrieve, transform and transmit information. The way the mind deals with each of these basic types of information processing provides relevant constrains for the design of visual representations. For instance, humans often, perhaps even typically, encode concepts in the form of hierarchies of schemas, which are information structures that coordinate attributes that describe and differentiate classes of concepts. These hierarchies of schemas underpin our ability to efficiently generalize or specialize concepts. Hence, we can use this knowledge to consider whether particular forms of symbolic correspondence will assist or hinder the forms of inference that we hope the user of the representation may make. For example, are the main symbolic correspondences in a visual representation consistent with the key attributes of the schemas for the concepts being considered?

Second, it may be useful for the designer to consider the broader nature of the tasks that the user may wish to do with the designed representation. Resource allocation, optimization, calculating quantities, inferences about of possible outcomes, classification, reasoning about extreme or special cases, and debugging: these are just a few of the many possibilities. These tasks are more generic than the information-oriented options considered in the 'design uses' column of Figure 27 in the article. They are worth addressing, because they provide constraints for the initial stages of representation design, by narrowing the search for what are likely to be effective correspondences to adopt. For example, if taxonomic classification is important, then separation and layering will be important correspondences; whereas magnitude calculations may demand scale mapping, Euclidian and metrical correspondences.

The third aspect concerns situations in which the visual representation must support not just a single task, but many diverse tasks. For example, a visual representation to help students learn about electricity will be used to explain the topology of circuits, make computations with electrical quantities, provide explanations of circuit behaviour (in terms of formal algebraic models and as qualitative causal models), facilitate fault finding or trouble shooting, among other activities. The creation of novel representations in such circumstances is perhaps one of the most challenging for designers. So, what knowledge can help? In this case, I advocate attempting to design representations on the basis of an analysis of the underlying conceptual structure of the knowledge of the target domain. Why? Because the nature of the knowledge is invariant across different classes of task. For example, for problem solving and learning of electricity, all the tasks depend upon the common fundamental conceptual structures of the domain that knit together the laws governing the physical properties of electricity and circuit topology. Hence, a representation that makes these concepts readily available through effective representation designed will probably be effective for a wide range of tasks.

In summary, it is desirable for the aspiring visual representation designer to consider symbolic correspondence, but I recommend they cast their net more widely for inspiration by learning about the human cognitive architecture, focusing on the nature of the task for which they are designing, and most critically thinking about the underlying conceptual structure of the knowledge of the target domain.

## 5.14 COMMENTARY BY BRAD A. MYERS

How to cite this commentary in your report

### Brad A. Myers



Brad A. Myers is a Professor in the Human-Computer Interaction Institute in the School of Computer Science at Carnegie Mellon University. He is an ACM Fellow, winner of six best paper awards, and a member of the CHI Academy, an honor bestowed on the principal leaders of the field. He is the principal investigator for the Natural Programming Project and the Pebbles Handheld Computer Proje...

Brad A. Myers
Brad A. Myers is a member of The Interaction Design Foundation

I have been teaching human-computer interaction to students with a wide range of backgrounds for many years. One of the most difficult areas for them to learn seems to be visual design. Students seem to quickly pick up rules like Nielsen's Heuristics for interaction (Nielsen & Molich, 1990), whereas the guidelines for visual design are much more subtle. Alan Blackwell's article presents many useful points, but a designer needs to know so much more! Whereas students can achieve competence at achieving Nielsen's "consistency and standards," for example, they struggle with selecting an appropriate representation for their information. And only a trained graphic designer is likely to be able to create an attractive and effective icon. Some people have a much better aesthetic sense, and can create much more beautiful and appropriate representations. A key goal of my introductory course, therefore, is to try to impart to the students how difficult it is to do visual design, and how wide the set of choices is. Studying the examples that Blackwell provides will give the reader a small start towards effective visual representations, but the path requires talent, study, and then iterative design and testing to evaluate and improve a design's success.

## References


- Nielsen, J., & Molich, R. (1990). *Heuristic evaluation of user interfaces*. Paper presented at the Proc. ACM CHI'90 Conf, Seattle, WA, 249-256.
- See also: http://www.useit.com/papers/heuristic/heuristic_list.html

## 5.10 BEHIND THE SCENES

*Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms below*

*Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.*

Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.

Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.

Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.


Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.

*Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.*

*Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.*



## 5.15 REFERENCES

Anderson, Michael, Meyer, Bernd and Olivier, Patrick (2002): *Diagrammatic Representation and Reasoning*. London, UK,

Bederson, Benjamin B. and Shneiderman, Ben (2003): *The Craft of Information Visualization: Readings and Reflections*. Morgan Kaufman Publishers

Bertin, Jacques (1967): *Semiology of Graphics: Diagrams, Networks, Maps (Sémiologie graphique: Les diagrammes - Les réseaux - Les cartes). English translation by W. J. Berg*. Madison, WI, USA, University of Wisconsin Press

Blackwell, Alan (2002): Psychological perspectives on diagrams and their users. In: Anderson, Michael, Meyer, Bernd and Olivier, Patrick (eds.). "Diagrammatic Representation and Reasoning". London, UK: pp. 109-123

Blackwell, Alan (2006): The reification of metaphor as a design tool. In ACM Transactions on Computer-Human Interaction, 13 (4) pp. 490-530

Blackwell, Alan and Engelhardt, Yuri (2002): A Meta-Taxonomy for Diagram Research. In: Anderson, Michael, Meyer, Bernd and Olivier, Patrick (eds.). "Diagrammatic Representation and Reasoning". London, UK: pp. 47-64

*Blackwell, Alan and Green, T. R. G. (2003): Notational Systems - The Cognitive Dimensions of Notations Framework. In: Carroll, John M. (ed.). "HCI Models, Theories, and Frameworks". San Francisco: Morgan Kaufman Publisherspp. 103-133*

*Carroll, John M. and Mazur, Sandra A. (1986): LisaLearning. In Computer, 19 (11) pp. 35-49*

*Engelhardt, Yuri (2002). The Language of Graphics. A framework for the analysis of syntax and meaning in maps, charts and diagrams (PhD Thesis). University of Amsterdam*

*Ferguson, Eugene S. (1994): Engineering and the Mind's Eye. MIT Press*

*Garland, Ken (1994): Mr. Beck's Underground Map. Capital Transport Publishing*

*Goodman, Nelson (1976): Languages of Art. Hackett Publishing Company*

*Gregory, Richard L. (1970): The Intelligent Eye. London, Weidenfeld and Nicolson*

*Horton, William (1994): The Icon Book: Visual Symbols for Computer Systems and Documentation. John Wiley and Sons*

*Ittelson, W. H. (1996): Visual perception of markings. In Psychonomic Bulletin & Review, 3 (2) pp. 171-187*

*Lowe, Ric (1992): Successful Instructional Diagram.*

*MacEachren, Alan M. (2004): How Maps Work: Representation, Visualization, and Design. The Guilford Press*

*Mccloud, Scott (1994): Understanding Comics: The Invisible Art. Harper Paperbacks*

*Micklewright, Keith (2005): Drawing: Mastering the Language of Visual Expression. Harry N. Abrams*

*Napoles, Veronica (1987): Corporate Identity Design.*

*Norman, Donald A. (1988): The Design of Everyday Things. New York, Doubleday*

*Petre, Marian (1995): Why Looking Isn't Always Seeing: Readership Skills and Graphical Programming. In Communications of the ACM, 38 (6) pp. 33-44*

*Pérez-Gómez, Alberto and Pelletier, Louise (1997): Architectural Representation and the Perspective Hinge. MIT Press*

*Resnick, Elizabeth (2003): Design for Communication: Conceptual Graphic Design Basics. Wiley*

*Richards, Clive (1984). Diagrammatics: an investigation aimed at providing a theoretical* *framework for studying diagrams and for establishing a taxonomy of their fundamental modes of graphic organization. Unpublished Phd Thesis. Royal College of Art, London, UK*

*Sellen, Abigail and Harper, Richard H. R. (2001): The Myth of the Paperless Office. MIT Press*

*Shneiderman, Ben and Plaisant, Catherine (2009): Designing the User Interface: Strategies for Effective Human-Computer Interaction (5th ed.). Addison-Wesley*

*Spence, Robert (2001): Information Visualization. Addison Wesley*

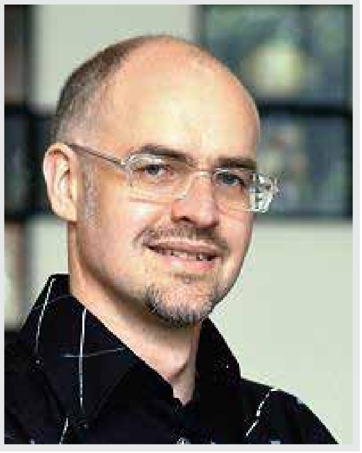
*Sutherland, Ivan E. (1963). Sketchpad, A Man-Machine Graphical Communication System. PhD Thesis at Massachusetts Institute of Technology, online version and editors' introduction by Alan Blackwell & K. Rodden. Technical Report 574. Cambridge University Computer Laboratory*

*Tufte, Edward R. (1983): The Visual Display of Quantitative Information. Cheshire, CT, Graphics Press*

*Ware, Colin (2004): Information Visualization: Perception for Design, 2nd Ed. San Francisco, Morgan Kaufman*

## About the author

### Alan Blackwell



© *Alan Blackwell*

Has also published under the name of:
"Alan F. Blackwell" and "A. Blackwell"

Personal Homepage:
http://www.cl.cam.ac.uk/~afb21/

Current place of employment:
University of Cambridge

I only have one big research question, but I attack it from a lot of different angles. The question is representation. How do people make, see and use things that carry meaning? The angles from which I attack my question include various ways in which representations are applied (including design processes, interacting with technology, computer programming, visualisation), various methods by which I collect research data (including controlled experiments, prototype construction, ethnographic observation), and the theoretical perspectives of various academic disciplines (including computer science, cognitive psychology, engineering, architecture, music, anthropology).

If you are based in Cambridge, you may like to attend the following talks on human-computer interaction.

This page lists a few large research themes and major projects illustrating them. Smaller projects, including contributions to research communities and research-related teaching, are described on my publications page, and other activities page.

## Crucible: Research in Interdisciplinary Design

Crucible is the Cambridge network for research in interdisciplinary design, which I founded with David Good. The network encompasses very many projects, funding sources and collaborators. Crucible projects include practical design work (as commercial consultants or in academic contexts) that draws on multiple disciplinary perspectives. We also carry out a significant amount of design research - investigating the processes of design work, developing facilitation processes for design activity, informing public policy related to the design of public value from academic research, and creating new and experimental software tools for designers to use. Many of these projects draw on my core expertise in visual representation.

## Software and Creativity

Many contemporary arts practitioners develop software, incorporate it into their work, or use software tools to extend their professional practice. This research theme involves collaboration with a wide range of artists, including many with international profiles, exploring the ways in which they use representations. We have created a wide range of new software tools and programming languages for composers, performers, choreographers, sculptors and others. Many of these projects are linked via the Crucible page.

## Social Media and Activism

Internet technology research is crucially dependent on understanding the social dynamics of the ways in which it is used, and collaboration with social scientists is essential to provide intellectual rigor and new insights. These projects have investigated the design and deployment of new social media, both in the world at large (various professional and political contexts), and within the University itself. In all cases, the representation of social relations around and within technical systems has been critical to understanding and the development of new understanding among all stakeholders. Many of these projects are linked via the Crucible page.

## Energy Monitoring and Usage

Electricity is invisible, so our awareness of environmental impacts arising from energy use is solely dependent on the quality of the visual representations provided of energy use. Ever since contributing to the design of the first generation of semi-smart domestic gas meters in 1991, I have taken an interest in the user interface of home energy controls and monitoring. Several of our projects are concerned with helping people understand and control the patterns of energy usage in their homes. Many of these projects are linked via the Crucible page.

## PhD Students' Research

- Michal Kosinski - psychological instruments for the assessment of business value in social networking technologies.
- Mo Syed - design techniques for incorporating social factors in technology development.
- Luke Church - social, cognitive, philosophical, artistic and technical perspectives on the manipulation of information.
- Chris Nash - Supporting Virtuosity and Flow in Computer Music.

- Cecily Morrison - Bodies-in-Space: investigating technology usage in co-present group interaction (thesis). Now a Research Associate in the Cambridge Engineering Design Centre.
- 29 Jan 2011: Added

## Your notes and thoughts on chapter 5

Record your notes and thoughts on this chapter. If you want to share these thoughts with others online, go to the bottom of the page at: http://www.interaction-design.org/encyclopedia/visual_representation.html

**NOTES:**

# Chapter 6

## Industrial Design

### *by Kees Overbeeke and Caroline Hummels.*

In loving memory of Prof. Dr. Kees Overbeeke *July 18th, 1952 - October 8th, 2011*

Kees left us unexpectedly on October 8th 2011, after a lifetime of dedication and warmth towards all the people who surrounded him. He was a passionate man. Kees was inspired, inspiring, engaging, dedicated, provocative and direct. He did not like easy and he was not easy, but he was a true friend and a great academic. He was a man who dared to dream and act upon it, and he encouraged others to do the same.

This chapter is one of his last publications. He started to write it spring 2011 and he asked me to join him a few months later. Although we weren't able to finish this chapter together, his legacy, dreams and beliefs are elucidated in this chapter. His memory will live on in our hearts and in our hands, as well as in those of our students, alumni, friends, family and all his national and international colleagues. And for those who do not know Kees yet, be inspired by his dreams.

*Had I the heavens' embroidered cloths,*
*Enwrought with golden and silver light,*

*The blue and the dim and the dark cloths*
*Of night and light and the half-light,*
*I would spread the cloths under your feet:*
*But I, being poor, have only my dreams;*
*I have spread my dreams under your feet;*
*Tread softly because you tread on my dreams.*

--- *William Butler Yeats*

Industrial design (ID) as a profession derives from the more general discipline of design. Design is a very old profession it simply refers to the way in which craftsmen have been designing and creating objects for millenia. Industrial design, however, is specifically seen as a product of the (first) Industrial Revolution when a single craftsman could no longer be responsible for every stage of the development of a product, from conception to sale, although the division between the creation of the product, the idea, and the manufacturing process had already appeared by the beginning of the sixteenth century when workshops specialised and trade expanded (Heskett, 1980). The use and development of new technology, production techniques and materials in the eighteenth and nineteenth centuries, however, enabled the development of mass production, standardisation, modularity and diversification of designs for new target groups. This development also required new forms of collaboration between different experts and new business models to stimulate mass consumption (Forty, 1986).

Although many books and studies on trying to capture the essence of design have been written, by authors such as Donald Schön (1983), Cross et al (1996), Bryan Lawson (2005) and Nigel Cross (2011), the field lacks a general definition of industrial design, or indeed of design in general. As Bryan Lawson and Kees Dorst (2009) state:

“ *One of the difficulties in understanding design, is its multifaceted nature. There is no one single way at looking at design that captures the ‘essence’ without missing some other salient aspects.* ”

*--Lawson and Dorst (2009)*

Industrial design can, for example, be seen as creating tangible propositions for the mutual benefit of both user and manufacturer; as creating design solutions for a broad market by integrating aspects such as form, usability, technology and business into a coherent whole; as problem finding, making sense and developing something to a preferred state; or as a mixture of making, thinking, contextualising and envisioning. Despite the lack of one coherent definition, the profession is well established, and all over the world industrial designers are members of professional ID associations.

This chapter on industrial design consists of four parts. The first part gives a short history of industrial design from the Industrial Revolution until the present day. The second part explains the essence of industrial design: the integration of theory and practice. The third part elucidates theories for industrial design such as Gibson’s theory of perception and phenomenology, theories which are now recognised and embedded in the design profession. The fourth and final part shows a set of “principles” about design and design research that sum up our views on, and accumulated experience with, design research.

## 6.1 A SHORT HISTORY OF INDUSTRIAL DESIGN

Industrial design surfaced in different ways at different places.

### 6.1.1 The Industrial Revolution in the UK

At the beginning of the Industrial Revolution, mechanised fabrication in the UK was still combined with individual craftsmanship and aimed at the continuity of societal and economic structures, as seen in Wedgwood tableware and the development of railway engines. The speed of technological advance throughout the nineteenth century, however, enabled manufacturers to seek greater profits by producing more for an expanding market with new customers. Manufacturers increased the degree and type of decoration on their products to enhance the status of their customers. This resulted in a variety of new styles such as Art Nouveau and Neo-Gothic. As a counterweight, the Arts and Crafts movement, which emerged in the 1860s and 1870s under the influence of artists such as William Morris, aimed at designing products that reflected the old ideals of craftsmanship (Forty, 1986; Heskett, 1980, Sparke, 1986).

Figure 6.1: Art Nouveau in the UK: A Tudric Pewter Clock designed by Archibald Knox, 1902-05.

## 6.1.2 The Industrial Revolution in the USA

During the Great Exhibition held in Britain in 1851 to celebrate world industry, Europe became acquainted with American products and developments, which aimed at modern industrial mass production and functionalism. Over the next half-century the 'American system' was not only characterised by its production techniques but also by the entire organisation, influencing the business processes as well as the functionality and appearance of the products, resulting in new products such as Remington typewriters, Singer sewing machines, Kodak cameras, McGormick reaping machines and Ford automobiles (Heskett, 1980). Due to increasing labour costs and a shortage of cheap labour, the USA focused on standardisation. In the early years of the twentieth century companies such as Henry Ford produced relatively low-cost cars that existed of assembled complicated mechanical parts. Concepts such as efficiency, standardisation and functionality became popular at the beginning of the twentieth century (Sparke, 1986).

Figure 6.2: Typewriter invented by Christopher Latham Sholes, Carlos Glidden and Samuel W. Soule between 1868 and 1873, and manufactured in 1874 by Eliphalet Remington & Sons Co as the 'Sholes & Glidden Type-Writer'. The colourful flowery decorations may have been intended to appeal to women, who were entering the typist profession.

### 6.1.3 Birth of the industrial designer

Although industrial design was commonly incorporated into industry by the end of the nineteenth century, the profession of the industrial designer was still rather ill-defined, meaning that the activities of artists, architects, craftsmen, inventors, engineers, technicians and other personnel of larger companies were all labelled as industrial design. Only at the beginning of the twentieth century did the legitimacy of the industrial designer surface as a person who integrated all of these activities, incorporating, e.g. technological, functional, aesthetic and business aspects (Sparke, 1986).

### 6.1.4 Modernism

In Germany, the Deutscher Werkbund, a precursor to Bauhaus, was founded in 1907, and aimed at integrating traditional crafts and industrial mass-production techniques. It focused on the societal role of design and art. Technology and machines could be used to improve people's taste and develop their cultural aspiration towards, for example, harmony and societal decorum. The benefits of technology were no longer available only for the elite. The movements arising from the turn of the century up to the 1930s, Bauhaus, De Stijl, Constructivism and Purism, advocated a new universal and objective style: the aesthetics of the machine. By glorifying machines and technical progress through objective shapes, they aimed at an improvement of the quality of life for all users. Modernism flourished (Heskett, 1980; Sparke, 1986).

As of the 1960s, the German company Braun, along with Dieter Rams, applied to products the Bauhaus principles of the earlier years of the 20$^{th}$ century, such as 'form follows function'. Rams deleted every superfluous detail and ordered the essential elements to give optimum support to functionality. He aimed for a neutral and harmonic aesthetic quality in order to allow the user to create his own 'image' of the product. As Rams mentions in his 'ten principles': "Good design is as little design as possible" (Bürdek, 2005).

Figure 6.3: Telephone ‘Frankfurt’ produced from 1928 by Fuld & Co., also known as the ‘Bauhaus telephone’. The shell and handset of the phone was designed by Marcel Breuer, the rest probably by Richard Schadewell.

*Courtesy of Christos Vittoratos. Copyright: CC-Att-SA-3 (Creative Commons Attribution-ShareAlike 3.0).*

### 6.1.5 Streamlining

At the same time, the idolisation of technology, progress and modernity was further effectuated from the 1940s until the beginning of the 1960s by styling products on the basis of symbols of progress, such as cars, aeroplanes and even satellites. Streamlining, originally derived from the shape of a drop of water, was ‘the’ way to express speed as a metaphor for energy that could liberate the user. It was the opposite of machine purity, and it diverted the attention from the inner workings of the product to its appearance. Thomas Hine (1986) coined the synthetic word ‘populuxe’ for this era of popular luxury for all.

Figure 6.4: A Googie (Populuxe) ashtray from about 1950 that does not only serve its purpose, but has risen to the level of decorative art. It was designed by the industrial designer Maurice Ascalon and manufactured by the Pal-Bell Company.



### 6.1.6 Electronic products

After the fifties, Modernism faded out in the West (Kint, 2001; Kint et al, 2010). Rationalism, objectivity, universalism and the application of technology and science to human needs and necessities were substituted by replaceability and consumerism. Without unifying ideologies, design got lost in ‘prettiness’. Designers were focusing more and more on the ‘package’ and appearance of the product. Innovation became more popular than invention, and the professions of the designer and the engineer grew further and further apart, especially during the last decades of the twentieth century (Sparke, 1986). The development of technology intensified; the size of the microchip decreased and, simultaneously, its possibilities increased. The technology push spurred on the functionality of appliances, thus offering the user unlimited possibilities and an enormous supply of electronic and digital products. Consumption and personal self-fulfilment were strengthened, and the gap between machinery and commodity increased (Hummels, 2000).

Industrial design missed out on this major industrial development; designers missed the electronic boat, more specifically ‘the interaction with the ungraspable’. The design of the physical form and the design of the interface of interactive products were separated. People working within the Human Computer Interaction (HCI) community such as usability engineers did the engineering and interaction thinking, and industrial designers were invited to beautify the new machines (Frens, 2006). They did this by expressing power, rationality, functionality and self-interest. The products were based on ‘cognitive’ interaction with displays and dozens of neatly organised buttons, as can be seen in, for example, the microwaves, telephones, business equipment, medical equipment, computers and photo and video equipment from the 1980s and 1990s.


Figure 6.5: A Grundig SVR format video recorder from about 1980.

### 6.1.7 Towards emotion and experience

In the 1980s, the post-modernist movement Il Nuovo Design, including the companies Studio Alchymia and Memphis, criticised this focus and approach. They advocated concepts such as diversity, discontinuity, eclecticism, ornaments, colour and experience in order to create a more enjoyable and fancy world (Horn, 1985). The adage 'less is more' was replaced by 'less is a bore'. Around the turn of the century, a weakened version of their ideas rapidly caught on in commercial appliances, such as Alessi kitchen appliances, Swatch watches and Apple's colourful iMac, and quickly seemed to turn towards a superficial form of fun and emotions, i.e. 'funnying' the look of products, which we could call 'form follows frivolity' (Hummels, 2000). Around the same time Bauhaus-like objects were reproduced for the elite, becoming terribly expensive and sold in 'design boutiques'. Architects were assembling workers' living units into skyscrapers in the US. Design became exclusive.

The shift towards fun, emotion and experience was also a result of the socio-cultural setting and the economic strategies of companies to survive. Companies introduced lifestyle brands to compensate for the loss of ideals, identity, belief systems and cultural references after the decay of Modernism. By adopting a brand lifestyle of one's choice, people found a way to regain a sense of identity (Brand and Rocchi, 2011). Designers helped to fuel consumption by repeatedly renewing the style of products, of which Swatch is a clear example. And they focused on brands and brand identity by designing the complete package, including machinery and commodity, appearance, services, 'experience', points of sale, advertisements, etc., such as Nespresso and Apple. Designers tried to envision the future through novel concepts, such as Philips' 'Vision of the Future' (1996), although many of those design concepts took the underlying computer logic and cognitive structures as given.


Figure 6.6: Funnying: Swatch Flik Flak Fifa World Cup Spain, 1982 The watch shows the time 2:22:16. The lower wristband says 'Looney Toons Active!' and 'Espana', and the wrist bands show Bugs Bunny, Daffy Duck, and the Tasmanian Devil playing soccer.

### 6.1.8 Knowledge economy

With the move into the twenty-first century, we are shifting from an experience economy to a knowledge economy. People no longer have to rely on brands and product lifestyle to signify and create their identity; instead they are nowadays able to create their own identity or identities through a variety of social platforms, such as Facebook and Twitter (Brand and Rocchi, 2011). We are becoming more and more digital and networked, which influences the profession of the industrial designer.

### 6.1.9 Reuniting HCI and industrial design

Today we see that the two worlds of HCI and industrial design are coming together again. HCI people move towards experience, and industrial designers embrace the digital, although one can clearly see their history and thus differences in approach and focus. One sees many HCI people and computer scientists who are seeking to bridge engineering and art, science and art, or the three, as can be seen in Robin Baker's book *Designing the future* (Baker 1993), or in Hiroshi Ishii's work at MIT MediaLab (http://tangible.media.mit.edu/). Moreover, the digital is often a starting point to connect to the real, e.g. with Ishii's Radical Atoms, and a cognitive approach towards experience is still dominant. One generally sees the engineering paradigm in this way of working and thinking based on externalising knowledge, changing the world and taking the machine perspective (Bartneck and Rauterberg, 2007), although the latter is more and more combined with a human perspective.

Industrial designers, on the other hand, tend to base their way of working and thinking on internal knowledge (often referred to as intuition), on changing the world and on a human perspective (Bartneck and Rauterberg, 2007). Contemporary industrial designers are trying to find a way to connect the possibilities of new technology, intelligence and social platforms to people's being in a physical and social world; to map the discreteness of the digital to the continuity of us being in the world. Moreover, industrial designers are exploring their new role in the upcoming transformation economy paradigm in which value is created in communities, by addressing societal issues together on a local scale with all the stakeholders involved (Brand and Rocchi, 2011). For example, RED, a 'do tank' that uses innovative design to tackle social and economic issues, which was set up by the British Design Council in 2004, is one of the initiatives to find new ways (Burns et al, 2006).

### 6.1.10 Reflections on this short history of industrial design

This short introduction to industrial design may give rise to the misconception that industrial design is only concerned with beauty, especially since traditionally design is taught in art schools (RCA, Design Academy Eindhoven, Design School CMU, Domus Academy). However, this is not the case. From the very early days, industrial design has been involved in emancipatory movements. For example, the German school Bauhaus, which operated from 1919 to 1933, was part of the socialist ideal of creating a 'new man' by ameliorating his environment. The slums should be replaced by houses that let in air and light; the furniture should be easy to produce and cheap to buy by using the latest in production techniques. The same went for cutlery and china. Beauty was an essential part of this endeavour, but not its aim. Industrial design gave direction to cultural developments. Beauty is defined in a context of transformation.

The reader might also have the misconception that all industrial designers are alike in their approach and focus. Although the education of industrial designers has strong similarities all over the world, at least two distinct approaches have arisen in design schools over the last half century. In both approaches young people were and still are trained for a few years in a design school to enter the profession. Mostly this is done, literally, according to the master-bachelor model. The master teaches the pupil by example in studios and workshops. At the end, the pupil produces a 'masterpiece' as a proof of the acquired skills and insights.

Based on this master-bachelor model, one approach emphasised the designer's artistic calling and withdrew into conceptual thinking based on 'intuition'. Another approach looked at science and wanted to become scientific, based on 'rationalism'. The problem is, however, that there are very few designers who are interested in science. They are professionals: why bother about science? This is of course too general a statement, but practice shows that these two worlds are often difficult to merge, although new design schools that aim at merging these directions have arisen and are currently arising, such as the department of Industrial Design at the Eindhoven University of Technology.

In the remainder of this chapter, we briefly sketch the value of integrating design theory and practice for the industrial design profession. We elucidate a few theories from psychology, philosophy and learning which have entered the design profession. We conclude the chapter with our own principles regarding industrial / interaction / intelligent design that are based on these theories, and we illustrate these principles with several design projects.

## 6.2 THE ESSENCE OF INDUSTRIAL DESIGN: INTEGRATION OF THEORY AND PRACTICE

One of the authors of this chapter, Kees Overbeeke, came from a scientific approach to industrial design and was not a designer.

"*My first contact with design taught me a lot. A designer asked me, 'Do you see that line in that car?' I just said, 'What line?' For me a line is something you draw with a pen on a surface. So, there was no line on the car.*"

As is evident, it may be difficult for a person from academia to understand the mindset of industrial designers. One of the reasons that industrial design, and the mindset of designers, is difficult to understand is that industrial design is essentially about *integration*.

### 6.2.1 Integration

Let us use an analogy to clarify what is meant by *integration*. Take medicine. Medicine is about the integration of knowledge from several established disciplines into practice. But no medical doctor is a specialist in any of the sciences involved. A family doctor has to know a lot, but essentially learns his profession by practising it. He has to acquire a 'feeling' for what may be wrong with the patient based on his knowledge and experience, and act accordingly. The same may be said of a designer. He has to know about materials, about beauty, about people, about culture, about business, about new technology, etc., and act accordingly. Again, the designer is no specialist in any of these fields, but knows how to combine them to create new opportunities.

There is one essential difference, though. The doctor tries to restore the old condition of the patient to make the patient better again. The designer creates something new, something that does not yet exist, and in doing so aims to make a better world. This is an important difference as the designer — as such — is not a 'problem solver'. Designers typically do not solve problems; they go one step beyond and create a new reality for people to explore. The designer as problem-solver is a popular misconception.

### 6.2.2 Knowledge that is relevant for practice

Now, how can science contribute to the profession of the industrial designer? There is a tendency to 'academise' all professions. That is a very good thing. Medicine really advanced once research into diseases took off. The scientific knowledge established by medical research centres is the direct reason that most of the readers of this text are alive today. At the same time, the family doctor has not lost his importance, nor has knowledge based on experience and intuition.

So, the challenge to academics is to establish a way to provide the practitioner with knowledge that he can effectively use in his practice while respecting his being-a-designer. There does not seem to be any correct or established way to do this, but we, the authors and their group, have enthusiastically tried for the last 25 years. We believe the criterion for success of people working on the research side of industrial design, i.e. academics, should be *relevance for professionals* (as opposed to relevance for other academics). Such criteria are, for example, 'How many designers use your methods?' or 'How many products using your insights are on the market?' We do not claim that we have the right answers, or that our approaches are the only ways of doing things. However, we have first-hand experience of practising and researching design as a discipline, which is for us the essence of industrial design: integration of theory and practice.

## 6.3 THEORETICAL UNDERPINNINGS FOR INDUSTRIAL DESIGN

Research cannot progress unless it is able to stand on the shoulders of giants. When we started our academic work, about 20 years ago, our aim was to regain control over the whole product design, its context and the way it enriched our lives as individual and social beings (for an overview see Overbeeke and Wensveen, 2003; Djajadiningrat et al, 2004; Hummels et al, 2007; Overbeeke, 2007; Hummels, 2012). We searched for basic principles and theories to unify our endeavour and found two theoretical frameworks from philosophy and psychology and several closely related theories that inspired us to develop the field of industrial design. We will explain these briefly in this part. Moreover, in Part Four we elucidate several other theories that are closely connected to the design projects we illustrate there.

### 6.3.1 Phenomenology, pragmatism and embodied interaction

The case for a philosophical underpinning of (interaction) design has been masterly made by Paul Dourish (2001) in his wide-ranging book about embodied interaction, starting from the European philosophical tradition. As we are not philosophers, we will be brief here. There is a long-standing tradition of dualism (object-subject) in the West. The best-known example is Descartes' ontology: " *je pense, donc je suis*." Being is thinking. The phenomenologists in Europe (e.g., Merleau-Ponty, 1958) and the pragmatists in the USA (e.g., Dewey, 1997) reacted against this dualism in the first half of the last century. Dualism seems to imply that the transcendent 'mind' is superior to the body, which is somehow ordered by the mind to act (Overbeeke et al, 2006).

Merleau-Ponty started from the fact that we have a lived (in the sense of experiential) and existential body. We do not have to prove that we exist. We cannot escape the fact. The body has a primacy: it is the way we are in the world. Or as Merleau-Ponty describes it: *"être au monde"*, which means not only being in the world but also belonging to it, having a relationship with it, interacting with it, perceiving it in all dimensions. Perceiving is an activity, and our body and skills are an inextricable part of our perception. We perceive the world in terms of what we can do with it, and by physically interacting with it, we access and express this meaning. Perception, through action, precedes cognition: reflection is a consequence of action. Moreover, we do not perceive ourselves as one more object in the world; we perceive ourselves as the point of view from which we perceive other objects (Merleau-Ponty, 1958; Trotto et al, 2011).

The fact that the body has inherence in the world, and that bodily experience pres our reflective capacities, is overlooked, even neglected, in most philosophies. That is why phenomenologists and pragmatists emphasise the, partly, embodied origin of knowledge, i.e., as reflection upon action. Donald Schön (1983) showed that this concept of 'reflection on and in action' was the core of design practice.

Theories like Embodied Interaction coined by Paul Dourish (2001) and related fields such as Tangible Interaction made the philosophy of phenomenology and pragmatism popular within interaction design. However, as soon as these theories were adopted by a large group of designers and HCI people, the subtlety and strength of Merleau-Ponty and Dewey's underpinnings was often strongly diluted, as is the case with 'affordances', a concept we will explain in the next section.

### 6.3.2 Ecological perception theory

Perception theory, within psychology, has always been a special field of study: it is ultimately about meaning. The field is strongly interwoven with design, e.g. through the impact of Gestalt theory on the Bauhaus Design School, and it is also strongly interwoven with HCI through mentalism and a cognitive approach to perception (Overbeeke et al, 2006).

We believe that one particular perception theory is of great value for design; one that is closely related to phenomenology. Gibson's theory of perception seems to be a natural ally for designers as it is a functional theory (Gibson, 1979). He is not only interested in *how* we perceive, but also at the same time in *why* we perceive. Gibson states that we perceive the world as essentially meaningful because we are 'fit' for the world through our action possibilities. We perceive the world in terms of what we can do with it, i.e., in terms of the action possibilities of our bodies; the functionality of the world reveals itself through manipulating the world, in interaction. This is useful for designers since meaning is put back into the level of action. And designers like that. It is very difficult to design for abstract meaning.

Two important concepts within Gibson's theory are affordances and effectivities. The essential notion of his theory is that we perceive the world in terms of what we can do with it. A person can perceive a surface as walkable (affordances) because he has feet and legs that can move (effectivities). For a fly, the ceiling is walkable because the fly has different effectivities, i.e., a different body. Similarly, a skater can perceive a surface or object as jumpable if he has the proper body, skills, speed and guts.

This approach, and more specifically affordances, has been introduced to design by Donald Norman (1988) and the present authors (Smets et al, 1988; Smets and Overbeeke, 1994). It now has a large following because it is about how we can design for doing, for acting. However, we do not agree with the misconception and simplification that an affordance is a physical characteristic of an object or the environment. We share Sanders' opinion that Gibson's brilliance was the unity of subject and object, which naturally includes one's intentions and every action that a specific organism is able to perform (Sanders, 1997; Michaels and Carello, 1981) .

Nor do we agree with the sometimes mechanistic interpretation of affordances. In 2000, Djajadiningrat, Overbeeke and Wensveen wrote:

" *We lament this clinical interpretation of affordance. People are not invited to act only because a design fits their physical measurements. They can also be attracted to act, even irresistibly so, through the expectation of beauty of interaction.* "

*-- Djajadiningrat et al, 2000*

### 6.3.3 Craftsmanship, reflective practice and constructivist learning

Related to and embedded in phenomenology, pragmatism and the ecological theory of perception, we find a variety of theories a few of which we briefly address here. Others can be found in Part Four of this chapter and in, e.g. Hummels and Frens (2011).

At the crossroads of design and phenomenology, one finds the concept of craftsmanship. Richard Sennett (2008) unravels this phenomenon in his book

*The Craftsman*, which gives a theoretical foundation and understanding of the core of professions such as design, including the designerly attitude and approach. In short, designing is about having a desire to do good work for its own sake. It is about trusting your senses, intuition, creativity, curiosity, imagination and skills to make (synthesise and concretise) and reciprocally to think (analyse and abstract). Designing is about localising (making a matter concrete), questioning (reflecting on its quality) and opening up (expanding its sense), and designers prefer ambiguity, uncertainty, open-endedness, complexity, resistance and they dare to fail. It is a vocation and an obsession (Sennett, 2008). Or one can say, designing is a way of living.

In terms of learning theories, we believe that designers develop and grow according to a learning model that follows an equilibrium — disequilibrium — re-equilibrium pattern (Piaget, 1971), where one goes from one stable state to another by way of disequilibrium, which is often chaos, through which one reaches order. Disequilibrium is the driving force of changing behaviour and development. Reflection and action are essential elements to regain order because they can change personal structures and ways of looking at the world and dealing with it (Doll, 1986). This fits Schön's reflective practice and Dewey's pragmatism, which are both based on the ability of professionals to know, reflect and learn in and through action; to learn by doing and through reflection gains an understanding that arises from experience (Schön, 1983; Dewey, 1997). So, designing is about jumping in, getting your hands dirty and learning from experience. Designing is rooted in a first person perspective and intermittently uses a third person perspective. Consequently, designers need to trust their intuition, use their common sense and dare to make mistakes, or as Schön puts it, by entering into an experience, without judgment, responding to surprises through reflection, we can learn from our actions. And as Merleau-Ponty (1962) states, perception, through action, precedes cognition: reflection is a consequence of action.

Due to the strong emphasis within design on learning by doing and reflection on action, a constructivist perspective on learning is prominent in the design field. The individual or cognitive variants of the constructivist paradigm assume the locus of knowledge construction to be in the individual learner; the social or situative variants assume this locus to be in socially organised networks (Birenbaum, 2003). Common to both perspectives, however, is the notion of activity: it is the learner who creates meaning, affected by and reflecting his socio-cultural environment. Constructivism is about learning and performing through practical application; that is, learning by doing, while simultaneously acquiring theoretical skills and building knowledge. Design programmes based on constructivism use the making skills of the designer as well as his analytical skills to gain knowledge (Hummels and Vinke, 2009). It is a unity of theory and practice, where experience plays a crucial role (Dewey 1938).

Design schools that use an approach to emphasise artistic calling and conceptual thinking based on 'intuition' generally fit in with a constructivist approach. Most design schools that moved towards science turned away from a constructivist approach, although one sees a change in new design schools such as our own department of Industrial Design at the Eindhoven University of Technology, which uses a self-directed, continuous and competency-based learning model (for more information, see Hummels and Vinke, 2009).

### 6.3.4 Pragmatic aesthetics and aesthetics of interaction

We conclude this discussion of theoretical underpinnings for industrial design with aesthetics, which have always been a core element of industrial design, especially the aesthetic aspects of the artefacts themselves. However, due to the rise in the development of interactive products, the aesthetics of interactive systems have shown a growing interest on top of the aesthetics of the artefacts themselves. Petersen et al (2004) point out that these two approaches to the aesthetics of design reflect Richard Shusterman's (2000) distinction between pragmatic aesthetics and analytical aesthetics respectively. From an analytic perspective, aesthetics arise as a product property. The focus of the design process here is on the aesthetics of appearance, on the creation of artefacts that are attractive and pleasurable to use. The pragmatic approach, on the other hand, is concerned with the aesthetics of use. According to this view, the aesthetics of an artefact emerge out of a dynamic interaction between a user and an interactive system resulting in what has been labelled 'aesthetic interaction' or 'resonant interaction' (Locher et al, 2010). According to Djajadiningrat et al (2007), different directions to aesthetics can be found within the pragmatic aesthetics approach, such as aesthetics of narrative (Dunne and Raby, 2001), aesthetics of actions (Buur et al, 2004), reactive graphics and computational aesthetics (Maeda, 1999) and semantics of movement of products (Kyffin et al, 2005).

Pragmatic aesthetics are gaining momentum nowadays with designers moving towards the digital and the HCI community moving towards experience, and many publications, including our own, have appeared in this field: Djajadiningrat et al (2007); Forlizzi and Battarbee (2004); Overbeeke et al (2002); Petersen et al (2004); Hummels and Overbeeke (2010) and Hummels (2007).

## 6.4 ACCUMULATED PRINCIPLES OF INDUSTRIAL DESIGN

Over the last 25 years, we, the authors, have been inspired by the philosophical views of the world described above. Consequently, we have formulated a set of 'principles' or 'beliefs' about design and design research that sum up our views on, and accumulated experiences with, design research. In this part, we explain these principles and illustrate them through examples.

1. Being in the world
2. The primacy of action
3. Reflection on action
4. Design methods
5. Intuition and common sense
6. First / third person perspective
7. Creating opportunities for transformation through subtlety

## 6.4.1 Principle 1. Being in the world

Industrial design is about people. It is about our lives, our hopes and dreams, our loneliness and joy, our sense of beauty and justice, about the social and the good. It is about being in the world.

### 6.4.1.1 Example: Ethics and aesthetics in intelligent product and system design

*Designed by: Philip Ross (PhD TU/e)*
*Project team: Kees Overbeeke, Loe Feijs, Stephan Wensveen, Tom Djajadiningrat and Caroline Hummels (TU/e)*
*Time frame: 2004-2008*
*Video: http://www.idemployee.id.tue.nl/p.r.ross/thesis/*

There is more to a product than its primary functionality alone. Take for example the mobile phone. It has a primary function. But it is more than just a product that opens an audio link to a person at another location. It offers an expression of a brand and a lifestyle. But it is also more than simply an object that expresses a 'lifestyle' through its styling. This work revolves around a third way of viewing the technologies we use — in terms of how they change our behaviour and the way we experience the world. See for example how the use of mobile phones has changed the way people manage their social relationships.

New technological developments allow for new kinds of change in behaviour and experience. Visions such as Ambient Intelligence sketch a future of intelligent products and systems that are embedded in the everyday environment: "weaving into everyday life until they are indistinguishable from it" (Weiser, 1991). Such omnipresent products and systems will bring about profound changes in our behaviour and experience. This new technological context gives industrial design the opportunity, but also the responsibility, to consider how to give these changes a desirable direction. This raises an ethical dimension of design of intelligent products and systems: what is a desirable change?

The work of Philip Ross (2008) investigates how to design Ambient Intelligent products and systems, taking into account this ethical dimension. His work explores how to design for the ethical dimension of products as well as for 'Aesthetic Interaction', which is a term that relates to beauty that encompasses the dynamics of human behaviour (see Part Three: Theoretical underpinning for industrial design). This is based on the Aristotelian notion of beauty, *to kalon*, viewing ethics and aesthetics as inextricably linked. So the question is how to design intelligent products and systems that invite people to change their behaviour in a specific way, through incorporating specific aesthetics in the interaction design. In other words, how can we make specific behaviour in the interaction with a product or system so attractive that people are enticed to engage in it?

Figure 6.7: Philip Ross using his intelligent reading lamp AEI.



Philip developed the intelligent lamp AEI to explore how to design a product dynamic form that has a specific effect at the person value level. AEI has three sets of behaviour targeting the values Helpful, Social Power and Creativity, i.e. the light pattern of AEI behaves and reacts to the user's actions in such a way that it aims at the user feeling helpful, creative or having social power (see these videos).

In order to be able to do this, the lamp has capacitive sensors placed beneath the porcelain outer shell to sense the position of the hand, and it has a small camera inside to detect and follow objects in front of the lamp. Moreover, it has an array of 48 high power LEDs that comprise the light actuator. AEI has been further developed over recent years and is currently commercially available as Fonckel, which has less emphasis on these three values, but is aiming at personalised, adaptive interaction to fit user initiated values (www.fonckel.com).

Figure 6.8: The intelligent reading lamp 'Fonckel' is commercially available.



Since *to kalon* in intelligent product and system design is fairly new, it also requires new methods and techniques, as will be further discussed in design principle 4 'design methods'. AEI could not have been designed in conjunction with similar design techniques such as 'Interaction Quality Framework' and the 'Perspectives model of behaviour in interaction' (Ross, 2008).

#### 6.4.1.2 Related themes and research: 'being in the world'

Philip Ross used the Human Value Theory of Shalom Schwartz (1992, 2004) to describe and measure people's ethical beliefs and the ethical implications of behaviour. During the last century, several disciplines emerged that studied the relationship between products and society, including ethical implications, such as the sociology of technology (Bijker and Law, 1992), the philosophy of technology (Heidegger, 1962; Achterhuis, 1997) and design history (Forty, 1986). Designers are not sociologists, philosophers or historians; they create and build intentional technology. We support Peter-Paul Verbeek's notion that "the ethics of engineering design should take more seriously the moral charge of technological products and rethink the moral responsibility of designers accordingly" (Verbeek, 2006). Therefore, we explain the concept of technological mediation to support designers taking this responsibility.

> " *Technological mediation concerns the role of technology in human action (conceived as the ways in which human beings are present in their world) and human experience (conceived as the ways in which their world is present to them).* "
>
> -- *Verbeek, 2006: p. 363*

As Verbeek states in the quote above, technology has a mediating role and influences how people interact. The implication is that designers are essentially 'doing ethics' implicitly or explicitly through their designs. So how can designers incorporate this dimension into their design process? Verbeek sees two options. The minimum scenario would be to assess if the design has undesirable mediating capacities and try to reduce or eliminate those. The second scenario would be to build in specific forms of mediation that are considered desirable.

Verbeek sees several mechanisms for technological mediation with its two perspectives: (human) action and experience, which we explain briefly.

### 6.4.1.2.1 TRANSFORMATION OF HUMAN ACTION

Human actions are not merely based on people's intentions or the social context in which they live, but also on the mutual influence of people and the material environment. Don Ihde (1990) refers in this case to 'intentionality' of a product and to 'multistability', which indicates that people can use and interpret products differently, depending on the context. James Gibson (1979) introduced the concept of 'affordances' which are context- and species-dependent action possibilities, and Bruno Latour (1992) introduced a concept from the film and theatre world; he uses the concept of 'scripts' to describe the influence of an artefact on human actions. For example, a ceramic coffee cup has the script to be washed after use, and a cardboard cup the script of being thrown away. Human action and behaviour have two mechanisms for transformation: invitation and inhibition. Products can enable and invite certain behaviour and inhibit other.

### 6.4.1.2.2 TRANSFORMATION OF EXPERIENCE

Don Ihde (1990) sees two relationships for mediating human experiences and interpretation of reality.

The first relationship is the embodied relationship in which the technological artefact becomes an extension of the human body. Heidegger (1927) calls this a 'ready-to-hand' tool that typically withdraws from the user's conscious attention (Coyne and Snodgrass, 1993). For example, when looking through a pair of glasses, one perceives the environment and not the glasses. The opposite of ready-to-hand is present-at-hand in which the product itself becomes the object of attention instead of the task you want to accomplish with the product. For example, a malfunction of a product, e.g. a hammer or a computer, may cause a focus shift from the task, e.g. hammering or browsing the internet to the actual product, e.g. 'What's wrong with it?' In Human-Computer Interaction, the state of being present-to-hand is often referred to as 'transparency of use,' i.e. that the product is so easy to use that its use becomes 'transparent.'

The second relationship is called the hermeneutic relationship in which a person needs to interpret (the information presented by) the designed artefact, as it is a representation of reality. For example, the thermostat *represents* the actual temperature, so one cannot *feel* the temperature through a thermostat. Nevertheless, the thermostat supports one in knowing what the temperature is and in regulating it.

When technology mediates our sensory relationship with reality, it transforms what we perceive by amplifying or reducing specific aspects. This transforming capability is called technological intentionality; technological artefacts are not neutral elements, but have intentions and play an active role in the relationship between humans and their world.

### 6.4.1.2.3 SENSIBLE ALTERNATIVE

The Sensible Alternative design from Jelle Stienstra clearly shows this mediating role of technology. When looking at contemporary smartphones like the iPhone, there seems to be a dominance of user interfaces based on menu structures and decision trees, albeit more graphic-oriented and gesture-based than traditional PC interfaces. Jelle Stienstra presents the Sensible Alternative, a concept that enables smart-phone users to navigate between applications by accessing action-possibility-dependent and personalised associated applications. A single added touch-sensitive spot on the underside of the smart-phone provides an alternative layer of interaction between man and machine, on top of hierarchical system architectures. Stienstra designed and prototyped this interaction layer that exploits the advantage of the continuous *and* the discrete powers of man and machine. In his case study, he explores several consequences of a phenomenological approach for designing complex systems, products and related services (Stienstra et al, 2011).

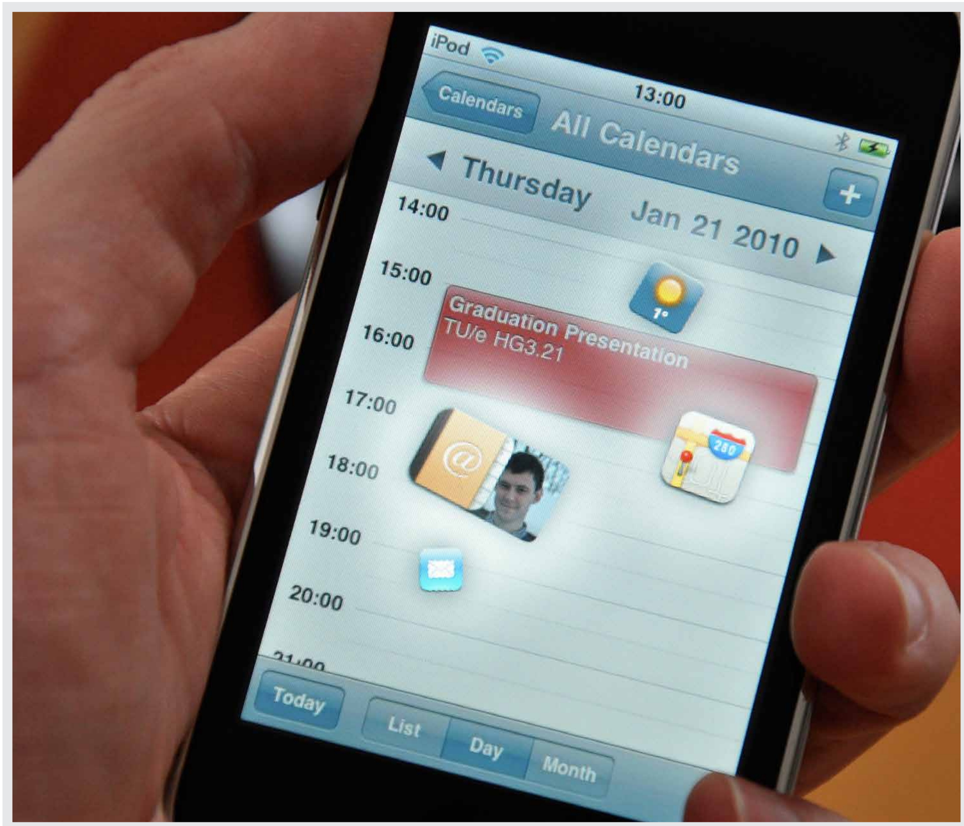



Figure 6.9: Sensible Alternative by Jelle Stienstra.

*Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.*

## 6.4.1.2.4 FROGGER AND INTERACTIVE MATERIALITY

Designed products transform people's behaviour, either through the sign character or the materiality of the product. Where often the static materiality of the products was inhibiting or inviting behaviour, interaction design can address the interactivity of the materiality and focus on the dynamics of the product. The exploration of the action- perception loop in a systematic and sensitive way using the Interaction Frogger framework (Wensveen et al, 2004) can inspire designers to transform behaviour in, and through, the interaction with products.

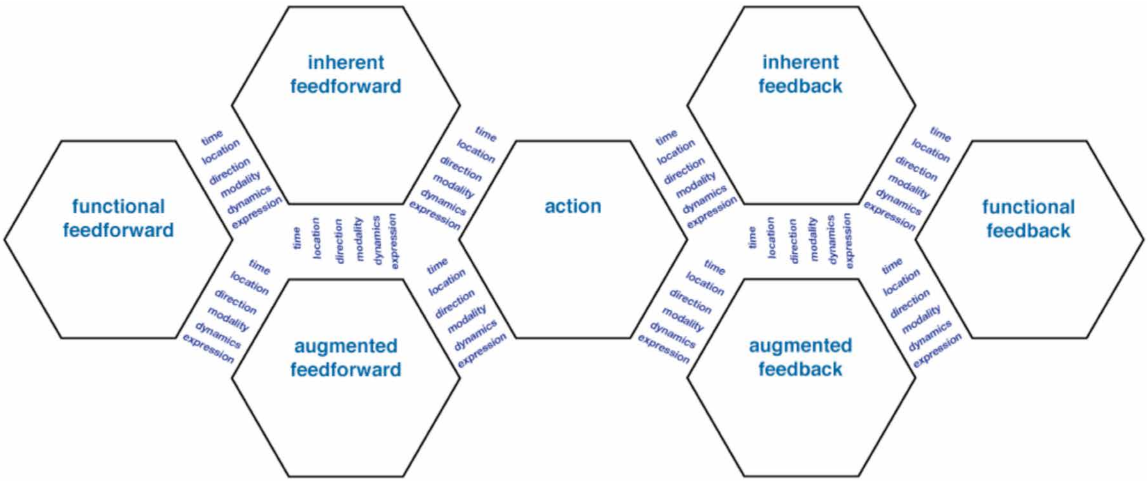



Figure 6.10: Interaction Frogger framework, showing all theoretically potential mappings between the action and the elements of perception (feedforward and feedback) (Stienstra, Bruns, Wensveen and Kuenen, 2012).



Designers can successfully transform behaviour on both a bodily and emotional level, by interweaving three essential steps in an iterative design process: an analysing step, a synthesis step and a detailing step. The analysing step focuses on gaining affirmation and appreciation by the designer for the current behaviour with a focus on the Interaction Frogger aspects of dynamics and expression. Synthesis focuses on the designing of the mapping within an action-perception loop. Within this step the ‘colouring’ for a targeted transformation of behaviour is tackled by exploring the six aspects of interactive materiality. Timing and scale related mappings can be applied over the input towards the output in order to amplify or reduce feedback contributing to feedforward and inviting transformation. The detailing step focuses on subtlety in interaction in order to have a ready-to-hand transformation of behaviour, exploiting the perceptual-motor and emotional skills. For this step, the designer should explore and seek boundaries within sensitivities of sensory systems respecting the uniqueness in skills and capabilities of the user. Designing for the transformation of behaviour, focusing on a person’s cognitive skills can now also be complemented with an approach that seems a better fit with third-wave HCI; the focus on the bodily and emotional aspects of interaction to transform behaviour (Stienstra et al, 2012).

### 6.4.2 Principle 2. The primacy of action

In accordance with Dourish's and others' approaches to epistemology, meaning cannot be detached from action. Meaning is in (inter)action. There is a primacy of embodiment.

#### 6.4.2.1 Example: Mustick

*Designed by: Joris Zaalberg, Tom van Bergen, Floor Mattheijssen, Wouter Kersteman (Former Master students TU/e)*
*Project team: Kees Overbeeke and Stephan Wensveen (TU/e)*
*Time frame: 2008*
*Video:* http://www.joriszaalberg.com/

The Mustick project is a very simple example of embodiment, but a very powerful one. v makes pre-recorded music play/stop. Once you have it in your hand, you start behaving expressively.

Conceived from a project on embodied interaction, Mustick is a design that allows non-musicians to perform songs of their favourite artists for family and friends at parties and social gatherings. The design uses the expressive powers of the human body to interact with pre-recorded music in meaningful ways. Mustick allows users to manipulate the playback of any song in real-time — simply by moving, shaking and swinging the device. The direct control of musical content creates a dialogue between user and music, turning the listener into a performer.


Figure 6.11: First version of Mustick.

*Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.*

Figure 6.12: The second wireless version of Mustick.

*Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.*

### 6.4.2.2 Example: DanceRail

*Designed by: Erik Toering, Pakwing Man and Frank de Jong (Former Master students TU/e)*
*Project team: Kees Overbeeke (TU/e), David Kirsch (UCSD), Philip Ross, Caroline Hummels (TU/e)*
*Time frame: 2009*
*Video:* http://www.dancerail.nl/

The DanceRail project involved a multidisciplinary team consisting of industrial design students from TU/e and Cognitive Science students from University of California, San Diego (UCSD). The goal of the interactive installation is to stimulate people to experience and think about movement with their body. The installation is related to an extensive research into dance creation at UCSD in February 2009 by Professor David Kirsch. His research concerns the new choreography of Wayne McGregor with his dance company Random Dance.

The DanceRail is an installation designed to engage your physical side — to encourage movement in a “dancerly” way. The shape of the rail was derived from an early version of dance phrases for a new choreography by Wayne McGregor. By running your hands over the rail, you will begin to move in non-everyday ways. The light is an enticement to explore and interact with the installation.


Figure 6.13: Roos van Berkel interacting with DanceRail.



#### 6.4.2.3 Related themes and research: 'the primacy of action'

Ever since the computer entered our daily and social life, it has ceased to be merely a means to perform our work and has become a means to help us pursue our lives. This development has connected the worlds of industrial design (ID) and human computer interaction (HCI), leading to the advent of fields like social computing, tangible interaction and embodied interaction.

Shaer and Hornecker (2010) give an extensive overview of the past, present and future directions of tangible user interfaces (TUIs), which interlink the digital and physical worlds. They explain how TUIs originated from Graspable User Interfaces (Fitzmaurice et al, 1995) and Tangible Bits and Bytes (Ishii and Ullmer, 1997). TUIs have become an established research area, and the word ‘tangible’ now appears in many calls for papers or conference session titles. Following diverse workshops related to tangible interfaces at different conferences, the first conference fully devoted to tangible interfaces and, more generally, tangible interaction, took place in 2007 in Baton Rouge, Louisiana. Since then, the annual TEI Conference (Tangible, Embedded and Embodied Interaction) serves as a focal point for a diverse community that consists of HCI researchers, technologists, product designers, artists and others (Shaer and Hornecker 2010).

Closely related to the field of tangible and embodied interaction is embodied cognition. David Kirsch, who, as described above, was deeply involved in the development of DanceRail, is strongly embedded in the field of embodied cognition (EC). Embodied Cognition merged in the early 1990s as an explicit reaction to the then standard cognitivist paradigm. Instead of a Cartesian split between mind and body, inner representations versus the outside world, embodied cognition can be found in the corner of Merleau-Ponty and Dewey, by appreciating the body’s special status. Within EC several variations can be found, such as distributed cognition and enactment (Dijk, 2012).

### 6.4.3 Principle 3. Reflection on action

A design theory must be a theory of action and the embodied in the first place, and of meaning in the second, and not the other way around. Reflection on action is the source of knowledge.

#### 6.4.3.1 Example: Sense6

*Designed by: Ivo de Boer, Joran van Aart, Bram Braat, Laurens Boer (former Master students TU/e)*
*Project team: DQI group (TU/e)*

*Time frame: 2008*

*Sponsor: Microsoft Research Design Expo*

*More info: http://dqi.id.tue.nl/docs/Eindhoven_designs_V1.pdf*

Sense[6] was developed for the Microsoft Research Design Expo, which aims to showcase exceptional design processes and ideas of a small number of universities from all over the world. The theme of the 2008 edition was 'Learning and Education'. The students developed Sense[6] during our six weeks Master Design Interaction Class, in which they proceeded from the brief to fully working prototypes using the reflective transformative design process (for the explanation of the RTD process, see related themes and research below). The project used 'reflection on action' on two levels: firstly, Sense[6] is a reflective practice platform for skateboarders to develop their skills; secondly, the design process of the design students depended heavily on reflection on action.

Figure 6.14: Sense6, a platform for learning extreme sports.

*Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.*

#### 6.4.3.1.1 INTRODUCTION

The Sense[6] concept explores non-obtrusive feedback on action through the connection between senses (synaesthesia) to support the 'learning by doing' paradigm. As explained in our theoretical underpinning, learning-by-doing is a key concept within Pragmatism. Sense[6] is a sharing platform for skateboarders, in which feedback on action is used not only to improve technique, but also to teach other people about their physical actions without the need to be there themselves.



Figure 6.15 A-B: A communicator with tokens is used to capture movements and share them online.

### 6.4.3.1.2 THREE ELEMENTS

Sense[6] consists of three main parts: 1) an online community, 2) a communicator with tokens and 3) sensor and actuator pads in grip tape and kneepads. The online community supports a platform for sharing skills over the internet. A skill is provided as a package of digital data that gives instruction through sensors and actuators. This data can be downloaded through the communicator to a token which contains the actual skill data.

The same communicator can be attached to the board to communicate the skill data from the token to the sensors and actuators embedded in the grip tape on the board and to the sensors and actuators in kneepads; it also provides the board with power. Light indicators and pressure sensors are embedded in the grip tape, and the kneepads contain bend sensors and vibrating actuators.

Figure 6.16: Light instructs a person to learn a variety of actions.



#### 6.4.3.1.3 SKILL DATA

The skill data provides information about location, balance and action rhythm while skateboarding. The position of the user's feet is first indicated through light. If this is correct, the user continues to the next step where pressure sensors measure the weight distribution and instruction is given by vibration. In the third step, a sound rhythm indicates the rhythm to be copied in the fourth step. Bending the knees and putting pressure on the board create this rhythm. This allows a very rapid movement to be learned by copying.

The skill data does not only apply to learning tricks, but can also be recorded onto a token, which can then be uploaded and exchanged through the community or exchanged in the physical world. By connecting two communicators together, one can copy data from one token to another through these communicators.

### 6.4.3.2 Related themes and research: 'the primacy of action'

In Part Three of this chapter, Theoretical underpinning of industrial design, we showed the importance of reflection for development. Phenomenology and pragmatism emphasise the, partly, embodied origin of knowledge, i.e., reflection upon action. Merleau-Ponty (1962) states that perception, through action, precedes cognition: reflection is a consequence of action. Professionals like designers have the ability to know, reflect and learn in and through action, to learn by doing, and through reflection gain an understanding that arises from experience (Schön, 1983, Dewey, 1997). Or, as Donald Schön (1983) states, by entering into an experience, without judgment, responding to surprises through reflection, we can learn from our actions. He also showed that this concept of 'reflection on and in action' was the core of design practice.

#### 6.4.3.2.1 REFLECTIVE TRANSFORMATIVE DESIGN (RTD) PROCESS

During these six weeks, the students positioned themselves in the theme of 'learning and education' by means of a transformative design vision in order to find their unique design challenges within a specific context. They went through several hands-on iteration cycles to create and deepen their insight into the design challenge. Moreover, they used co-reflective techniques to validate their proposals in real-world settings with users. In short, they used the Reflective Transformative Design process (RTD process). The RTD process is especially created to support designing disruptive innovative products and intelligent, open systems (Hummels and Frens, 2008; Hummels and Frens, 2011).

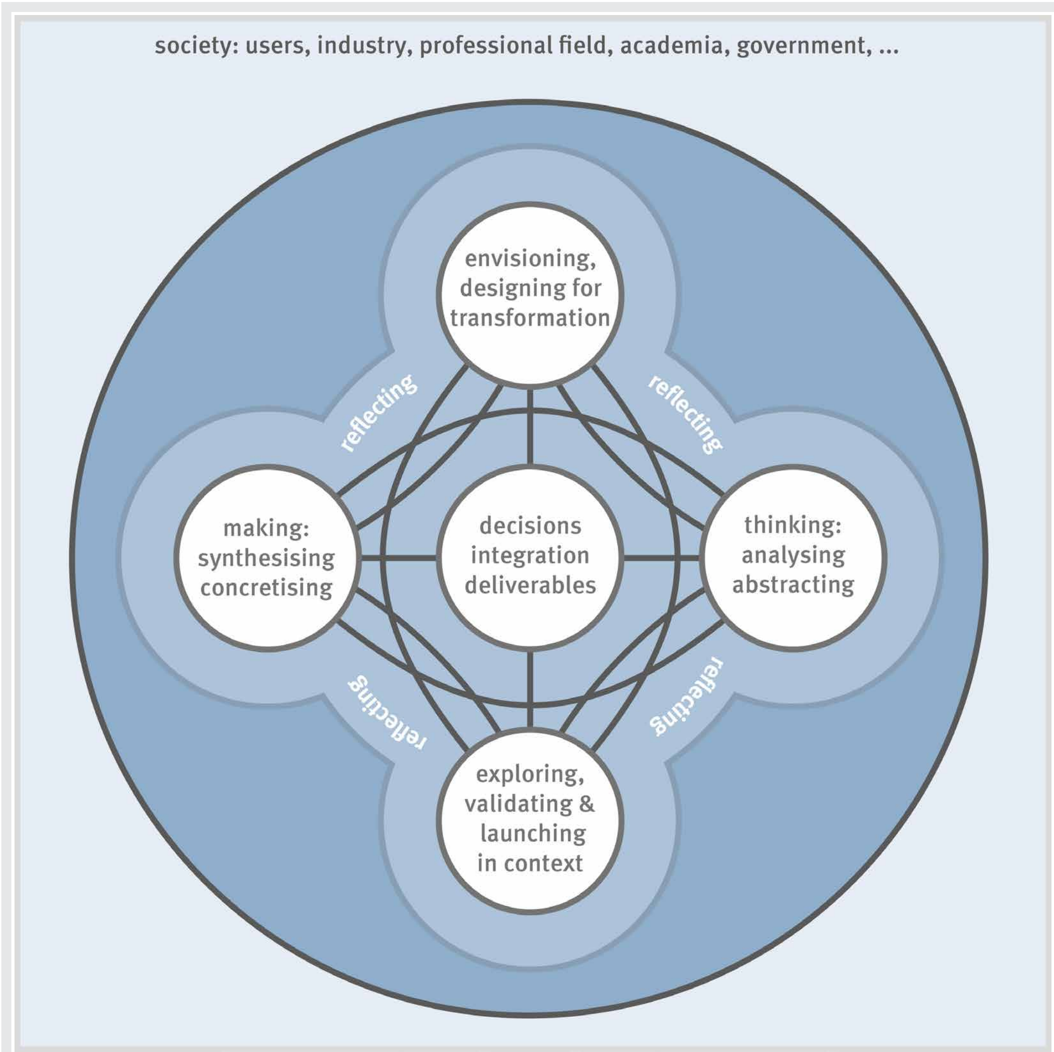



FIGURE 6.17: The Reflective Transformative Design process.



Developing design solutions, which is placed in the centre of this model, can be seen as a process of taking decisions based on too little information. These decisions are always conditional and can change over time based on the available information and actions taken.

The RTD process knows two axes: vertically we distinguish drives, and horizontally we distinguish strategies for information gathering to direct design decisions. The first drive is information gathering to direct the design decisions through the designer's vision (top circle), and the second drive is information gathering to explore and validate design decisions in the everyday living context of and with users, even beyond launching the system, product and services in the market (bottom circle). The drives are incorporated within two strategies that generate information and that reciprocally provide focus for each other. The first strategy revolves around design action, both synthesising and concretising, such as building an experienceable prototype (left circle). The second strategy revolves around academic thinking: analysis and abstraction (right circle).

Dependent on the person, context, or phase within the design process, designers determine where they start and the order of the activities. This way, the process supports flexibility, diversity and individuality, and it can even enhance chaos, moving from disequilibrium to re-equilibrium. The designers also determine how often they switch from one activity to another, although a high pace is recommended, especially during the early phases of the design process, but also during the later phases since this enables the designer to get a fast and good insight into what is happening within interaction in a diverse social context and market.

### 6.4.4 Principle 4. Design methods

Methods used in a design process must be rooted in design practice, in the socio-cultural and multi-cultural environment, invigorated by experimental and technological methods from other disciplines.

#### 6.4.4.1 Example: Light Through Culture

*Designs by: Master students (TU/e and University of Siena)*
*Project team: Kees Overbeeke (TU/e), Patrizia Marti (UoS), Rombout Frieling, Remco Magielse, Richard Appleby, Caroline Hummels (TU/e)*
*Time frame: 2011*
*Sponsors: University of Siena, Erasmus, ILI/OpenLight*


Figure 6.18: The pilgrimage route based on light in the Museum Complex Santa Maria della Scala.



Light Through Culture is an international design school which explores the theme of complexity in learning environments. The aim of the school is to weave the newest technologies and the rich existing culture into a new canvas for making and thinking. Learning is meant as a way to (re-)incorporate culture and making into thinking. The school is hosted by the Museum Complex Santa Maria della Scala in Siena, Italy.

The historical backdrop of the project was Via Francigena, the ancient pilgrimage route established in the Middle Ages from Canterbury to Rome, thus passing through England, France, Switzerland and Italy. Travelling along the route, the pilgrims stopped at the hospital of Santa Maria della Scala, where they could be given shelter and care on their way. Pilgrims started a long trip towards the enlightenment, the hope, the alleviation.

Being immersed in this rich historical-cultural context, the project participants were asked to design an experiential path in the underground museum where it intersects with the ancient Via Francigena. The design school tried to answer the question as if contemporary museum visitors were new pilgrims. This was done by exploring history and culture by means of innovative light technologies. Learning developed while building a real path in the underground museum. At the end of two weeks the pilgrimage experience was offered to hundreds of visitors, based on a route through four spaces (see film; Marti and Overbeeke, 2011):

Scala: “The road that seems easy might lead us nowhere.”

Finestra: “Life has a certain end but there is hope.”

Carnaio: “Life is cruel, how can we go on? The words comfort us.”

Lavatoio: “Life is purifying. Water will purify us, and we are innocent again.”

Opening the results of the design activity to the experience of real visitors and reflecting on how people feel, perceive and make sense of their experience is a strong means of learning to become a designer. Learning confronts you with the whole complexity of a real environment: the results of the school were not only texts, but also physical, virtual and mixed new realities consisting of new ways of presenting and adding new dimensionalities to the existing world.

#### 6.4.4.2 Related themes and research: ‘design methods’

Design researchers are continuously confronted with the challenge ‘ how to design for …’ Which process or method is appropriate for design research? Within the DQI group, we strongly believe in the strength of Research-through-Design (RtD) And since there is no uniform definition, we briefly explain what we mean by Research-through-Design.

Often RtD is associated with Bruce Archer's 'research through practice' (Archer, 1995), which can be seen as a process in which scientific knowledge is generated through consequent cycles of designing, building, and experimentally testing experiential prototypes in real-life settings. This implicates that RtD aims at studying an effect in a possible future, instead of understanding the world, as is the objective of traditional science (Stappers, 2007).

Designers are typically operating in a constantly changing context which can never be accurately modelled (Wakkary, 2005); thus a reductionist approach to addressing this context and situations within would fail (Zimmerman et al, 2007). Consequently, both designers and design researchers have to embrace the richness of a complex design situation, be fully immersed in the context of the case, and act designerly and in a way that is appropriate for the specifics of that situation. This implies developing detailed design prototypes that allow for this diversity, subtlety and richness during their confrontation with the world (Hengeveld, 2011; Stolterman, 2008). Therefore, RtD relies on the generation of wealthy, experienceable prototypes, and an evaluation *in situ* in a physical, human and experiential sense (Wakkary, 2005) will consequently result in equally contextual scientific results, leading more to conditional regularities than to general laws (Hummels, 2000). The prototype is the physical, experiential manifestation of this; the carrier of integrated, contextualised knowledge, the physicalisation of a design rationale. This means that when the test subject interacts with a prototype, he interacts with the designer-researcher's line of thought. Bart Hengeveld (2011) thus concludes that researcher and designer should be one and the same person.

Research-through-Design is also part of the book about design research that appeared in October 2011: *Design research through practice: From the lab, field, and showroom*, by Ilpo Koskinen, John Zimmerman, Thomas Binder, Johan Redström and Stephan Wensveen. The book gives a complete overview of how international design research is conducted in some major European and American universities. The authors took on the challenge to write an overview that is theoretically informed, rooted in practice, showing many examples and relevant to both research and practice.

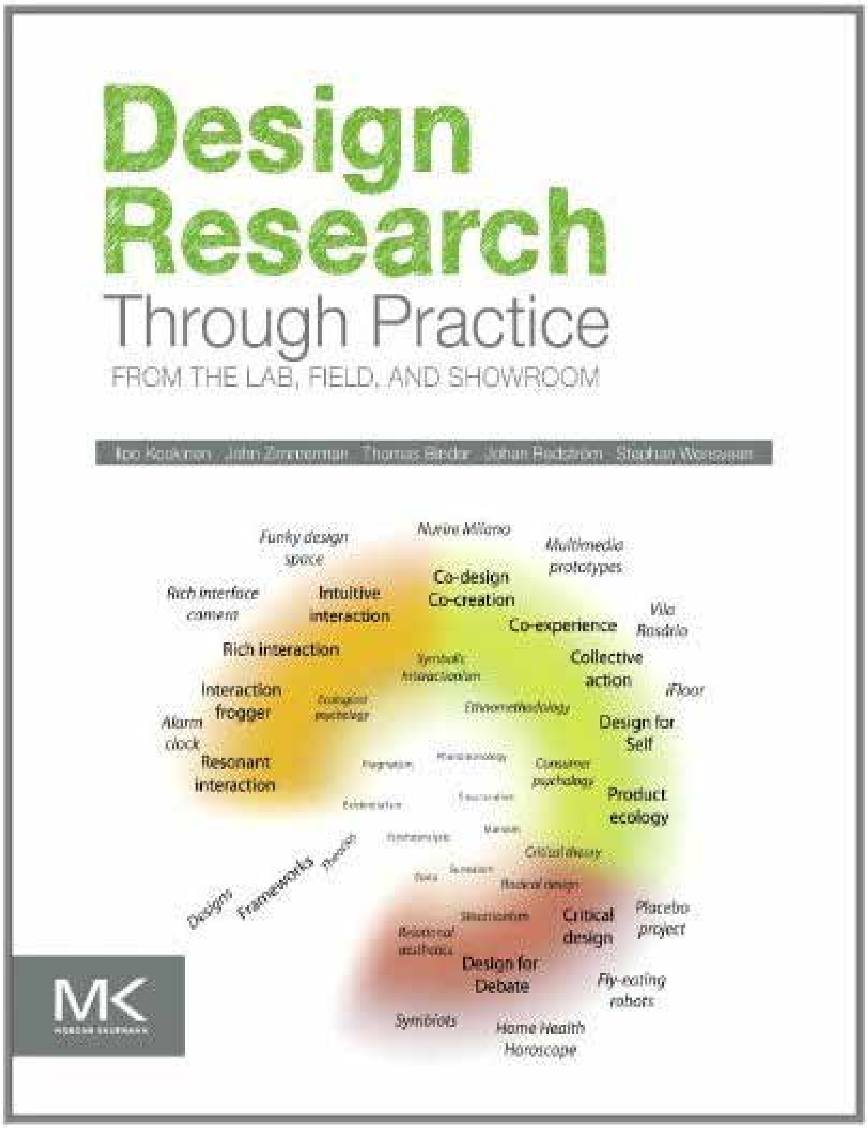



FIGURE 6.19: Book cover: Design research through practice.

### 6.4.5 Principle 5. Intuition and common sense

Intuition and common sense are paramount and should be leveraged to the maximum. "*Le sens commun n'est pas si commun*," Voltaire said, meaning common sense is not so common.

#### 6.4.5.1 Example: LinguaBytes

*Designed by: Bart Hengeveld (PhD TU/e) and Riny Voort (Viataal).*
*Project team: Jan de Moor (Radboud University) Hans van Balkom (Kentalis / PonteM), Caroline Hummels and Kees Overbeeke (TU/e)*
*Time frame: 2006-2011*
*Sponsors: Dr. W.M. Phelps-Stichting voor Spastici (main sponsor), Stichting VSB-Fonds, SKAN Fonds, Nederlandse Stichting voor het Gehandicapte Kind, Nationaal Revalidatie Fonds, Stichting Kinderpostzegels Nederland, Johanna Kinderfonds and Stichting Bio-Kinderrevalidatie*

*LinguaBytes* is an interactive play-and-learning system aimed at stimulating the language development of children with multiple disabilities including minimal or no verbal abilities between the ages of one and four years old (Hengeveld, 2011). The foundations of early language acquisition are laid by early parent-child interaction in the first years of a child's life (Snow and Ferguson, 1977; Tomasello, 2003). When early language development is distorted, as is generally the case with children with multiple disabilities, parent-child interaction does not start or progress normally. Not only does this cause impediments in the child's linguistic skills; it also has repercussions on perceptual-motor, cognitive and emotional skills since the developments of all skills are interdependent at this age.

Augmentative and alternative communication (AAC) is a term that refers to an interaction in which technology is used to enhance the communicative skills of handicapped individuals. It may seem obvious that products for this very young user group (aged one to four) require radically different interfaces than the ones we know from the PC, which is a machine originally designed for office use.

Figure 6.20: One of the potential user of LinguaBytes demonstrates the inadequacy of the PC, its input and interaction style.



Despite the commercial success of many AAC systems, the products have unfortunately not been given these radically different interfaces . The cognitive load of these systems is high, i.e. most AAC devices are not organised in ways that reflect how young children think (Coker and Shook 2006). Moreover, the current generation of AAC systems are not particularly appealing to young children in comparison with toys (Light et al, 2004). They resemble PCs in structure (menus and decision trees), input (mostly button-like) and output (often screen-based display). Despite the useful endeavour to develop a variety of special input devices for these youngsters, one wonders why young children are placed behind a desktop computer, which was originally designed for office work. Moreover, PCs are typically designed for exclusive use, i.e. one person sitting behind the screen with input devices, which is far from ideal to enhance interaction between parent/therapist and child.

In response to this situation, LinguaBytes challenges multi-handicapped children to capitalise not only on their cognitive and linguistic skills to develop their language and communicative skills, but also on their perceptual-motor and social skills. LinguaBytes is a modular system in which language is offered to children in a physical, playful form. Using a large collection of playful materials, children can read interactive stories and do linguistic exercises, preferably together with a caregiver, therapist or teacher.

LinguaBytes consists of several ‘spaces’. The first space is primarily controlled by the child, i.e. using the LinguaBytes play-and-learning materials to interact with stories and exercises. By placing the child in control, LinguaBytes contributes to the child’s sense of independence and thus stimulates self-esteem. In the second space, the carer has primary control over the choice and timing of stories and exercises and as such is responsible for giving structure to the child’s linguistic development. Additionally, there are two spaces where carer and child either have shared control or joint attention: both can communicate with LinguaBytes’ collection of play-and-learning materials and communicate about what they see happening in the interactive stories and exercises, shown on LinguaBytes’s output module.

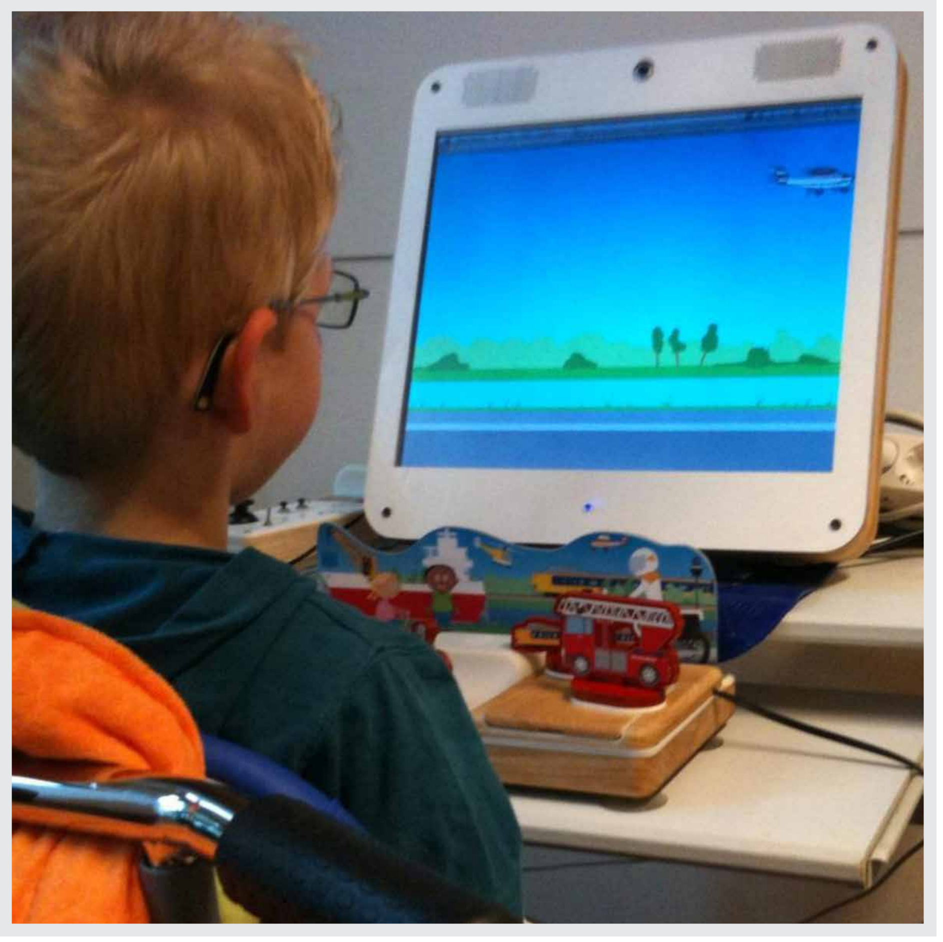


FIGURE 6.21: The LinguaBytes system with different exercises.


FIGURE 6.22: The LinguaBytes system with different exercises.



Intuition played an important role while designing LinguaBytes, for both the designer-researcher and for the children. The necessity to use this natural character of designers was even amplified because non- or hardly speaking children cannot be involved through regular research methods (e.g. interviews are impossible), but they can and do show the value LinguaBytes has for them through their behaviour. For example, repeated actions, laughing, taking over control and a subtle smile can all be hints to see if a child has understood and liked the exercise. Consequently, the designer, Bart Hengeveld, was often present at the rehabilitation centres to engage with the children and the therapists.

Since Bart had to trust his intuition, he went through a highly iterative research-through-design process, thus exploring a large scope of possibilities with a variety of toddlers and carers. As explained in the previous principle, making these large amounts of concepts and ideas experienceable enabled him to communicate with the children via his designs. And since the children and the therapists interacted with his physicalised design rationale, he could stay close to his intuition. Based on his years of experience as designer, e.g. of toys and interactive systems, his experience informed his intuition.

Nevertheless, he also developed creative methods to trigger his intuition in fields in which he was less experienced, such as the severe motor disabilities of some of these children. For example, by re-contextualising a situation, designers often get ideas that are not always useful, but highly inspiring. And as the obvious solutions will surface anyway, Bart preferred for this challenge to invest in the less obvious ones, which is again a way to trigger intuition. For example, he gave a group of the designers the following exercise, to explore the challenges of children whose movements are involuntary:

"*Oh dear, you are not in luck: the location where you organise a chess tournament lies at a notorious fault line. Every other minute the whole venue shakes so much that not only the chess pieces topple, but the players as well. It is too late to relocate or cancel. How can you prevent the chess games at the tournament from going to pieces?*"

*-- Hengeveld, 2011*

But intuition was also essential for the children interacting with LinguaBytes. Given the limitations of their communication skills, the children who use LinguaBytes can only use products that are intuitive and do not need manuals or instruction. LinguaBytes gives importance in two ways to the primacy of action and embodiment, while taking into consideration the perceptual-motor skills of the toddlers (e.g. the restricted motor skills of children with cerebral palsy). Firstly, the system invites certain behaviour: for example, the white trays of the exercise module afford placement of the (white base of the) inputs. The relation in size between tray and the height of the tray, the size of the inputs and their weight all support the child in placing the inputs in the tray. Finding this balance was an iterative process of building many different trays and inputs. Secondly, for children with severe motor limitations, the actual placing of an input is smoothened through subtle and unobtrusive support like physical restrictors such as magnets to 'guide' certain movements.

#### 6.4.5.2 Related themes and research: 'intuition and common sense'

The designer of the LinguaBytes project, Bart Hengeveld, opened up his intuition through his *making* skills. Firstly, by building a variety of experienceable prototypes, he enabled the toddlers and their carers to have access to and express meaning. In contrast, abstract ideas cannot be experienced or interacted with (perhaps only imaginatively). Secondly, *making* enabled him to explore the unknown by trusting his senses, exploring resistance and ambiguity and tapping into his intuition. Dijksterhuis and Nordgren (2006) show that intuition, or *unconscious thought* as they call it, is better suited for dealing with complex matters than conscious thought. Designing, which is based on creating, is the highest form of (cognitive) complexity, according to the Revised Bloom's Taxonomy (Anderson and Krathwohl, 2000).

Intuition is usually not considered as an 'official' *modus operandi* because using intuition does not contribute to making the process repeatable by others. Yet, because of the complexity of design processes and the intrinsic complexity of users, intuition is an indispensable component in design; it is the tool that empowers us to make choices in the iterations of a design process. Because "intuition begins with the sense that what is not yet could be" (Sennett, 2008, p. 201), it involves skills, as skills are our way to make sense of the world and transform it (see principles 2 and 3). Intuition is necessary to make leaps in the design process, and it is “an imaginative experience … that guides us towards what we sense is an unknown reality latent with possibility” (Sennett, 2008, p. 213). Therefore, training intuition is essential to become skilled in designing systems and products.

Sennett proposes four stages to enable intuitive leaps: reformatting, adjacency, wonder and gravity (Sennett, 2008, p. 209).

1. Reformatting is *the willingness to see if a tool or practice can be changed in use*; it is the abstraction phase. Can we reach the core of what a certain tool or practice is? For example, one of the authors, Caroline Hummels, likes ballroom dancing. When trying to reformat this ‘social dance in conventional rhythms’, it boils down for her to having the right balance between, on the one hand, letting go, being in flow, ‘stepping out of one’s body’ and floating, and, on the other hand, staying in charge, stretching the entire body, owning the space and being present.

2. In order for the intuitive leap to take place, the designer uses adjacency. *Adjacency occurs when two unlike domains are brought together*. Based on her impression of dancing, Caroline started doing sketching exercises with a pencil with the aim of achieving the same feeling as dancing. And she also tried to do it the other way around: drawing while dancing.

3. Thereupon tacit knowledge is dredged up into consciousness to do the comparing, which can result in surprise. As Sennett reminds us, the word ‘wonder’ in ancient Greek is embedded in *poiein*, a word that indicates the act of making, which is also the root-word for poetry. By cross-pollinating dancing and sketching, it turned out to Caroline’s surprise that dancing felt lighter when trying to draw the dance with her body…

4. The last stage of intuition which allows us to see beyond what is there, once again through the use of abstraction, is "recognising that a leap does not defy gravity; unresolved problems remain unresolved in the transfer of skills and practices" (Sennett, 2008, p. 211). Back with two feet on the ground. Although ballroom dancing feels lighter, and she is training twice a week for competitions, the physical limitations are there, and they will frame the progress in lightness.

### 6.4.6 Principle 6. First / third person perspective

Design practice and design research are powerful generators of knowledge. They are a way of looking at the world and transforming it. This way of looking is rooted in a first person perspective while intermittently taking a third person perspective. Consequently, design relies on connecting the sensorial, intuitive, to the abstract, analytical.

#### 6.4.6.1 Example: Master's Classes 'Designing for points of view' and 'Aesthetics of interaction'

*Designs by: Master students (TU/e and Tsukuba University Japan)*
*Lecturers: Caroline Hummels, Kees Overbeeke, Ambra Trotto and Michael Cruz-Restrepo (TU/e)*
*Time frame: 2006-2011*

In a course, students were asked to investigate deeply the 'design of meaning'. What is important for designers to realise is that meaning construction also occurs during the design process. As we do in all our work, we took phenomenology as a starting point, which implies that designers perceive themselves as the point of view from which they perceive systems and products. Consequently, they are a part of their designs. They are designing from a first person perspective while intermittently taking a third person perspective

(Trotto et al, 2011). Within the first person perspective, they are doing in action, reflecting in action and experiencing the situation. Simultaneously, they bring in their own value system and skills. Consequently, their designs will be meaningful for them in a different way from that of someone else. When stepping out towards a third person perspective, they can reflect upon their actions or upon the situation.

In line with the phenomenological approach, the students were asked to design an empowering/enabling tool that would allow a person to experience someone else's skills. To be able to design such a tool, the students went through the different steps of documenting and reflecting upon their own and each other's skills. One group, Yves Florack's, took up snowboarding.

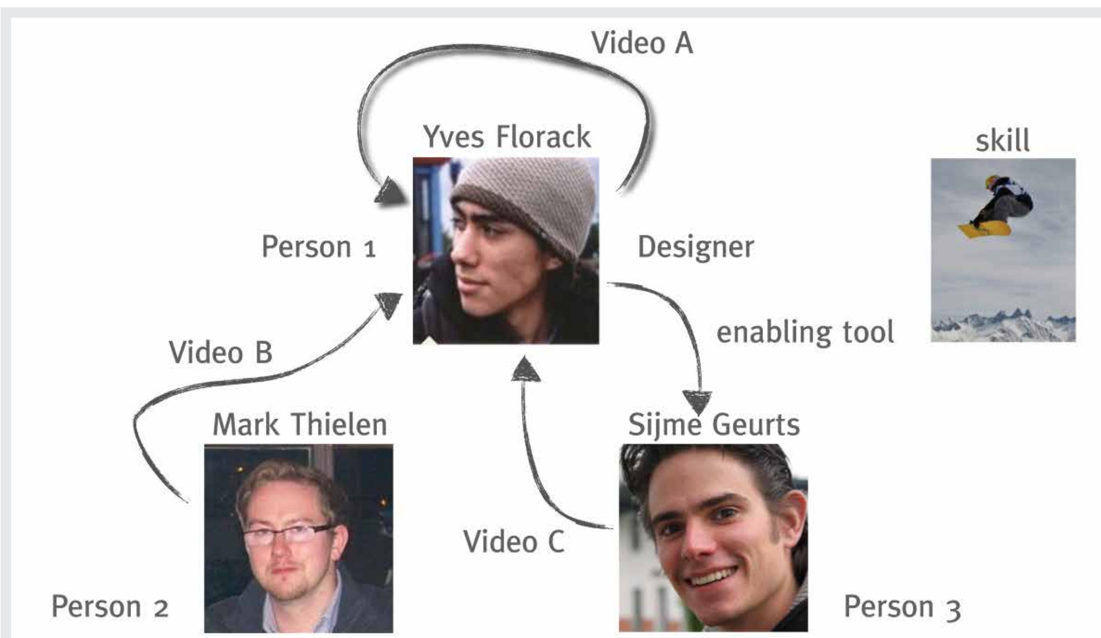



Figure 6.23: Set-up Master's class 'Designing for points of view'



- During the first day, Yves documented in a short video documentary his own skill: snowboarding (1st person perspective).

- During the second day, he tried to show and tell Mark at an indoor ski centre what snowboarding meant for him, and Mark translated his impression of Yves's skill into a short video documentary (2nd person perspective).
- Mark showed his video to Yves, who used it as a mirror to reflect on his own skill, and subsequently, Yves fine-tuned his first video.
- During the third and fourth day, Yves made a cutting tool for Sijme to experience the salient element of snowboarding: moulding and carving of a soft material by means of an artificial, fixed extension of the body. The soft material was in this case a sandwich, and moulding and carving were done with glove-based utensils.
- On the fourth day, Sijme was asked to learn to master the accompanying skill of making sandwiches in a time span of one hour, while being recorded on video.
- Thereupon, Sijme turned this video into a documentary by reflecting on the tool, skill and experience.
- During the final presentation on the fifth day, Yves reflected on the process, showing simultaneously videos A, B, and C and the design (third person perspective).

It was clearly seen and experienced by these three students that all perspectives and videos were different. Moreover, the students were encouraged afterwards to take a first or/and second person perspective more often during the design process, e.g. during interviews, which we indeed saw happen in later projects.





Figure 6.24 A-B: Sijme Geurts is trying to learn the skills of making sandwiches with Yves Florack's tool.

### 6.4.6.2 Example 2: Aesthetic of interaction

As one can see, by continuously moving from a 1st to 3rd person perspective and vice versa, design connects the sensorial and intuitive to the abstract and analytical. We explored this continuous switch and the strength of reflection on action within the one-week Master's class "Aesthetics of interaction" (Hummels et al, 2009).

We formed groups of three to four students and introduced them to pieces of classical music, ranging from renaissance to present-day music. The teams were given the overall assignment of designing an interactive product where the interaction had to have the same quality/experience as the provided music. We used music because it has reached a form of complexity of beauty which interaction design has not yet reached. Moreover, music helps open up the sensitivity of the students.

The students went through a design process consisting of five methods that utilised the balance between the sensorial, intuitive, and the abstract, analytical:

Figure 6.25: Films: Students listened to the music and then made a one-minute film showing the link between music & quality of interaction (relying on their senses).

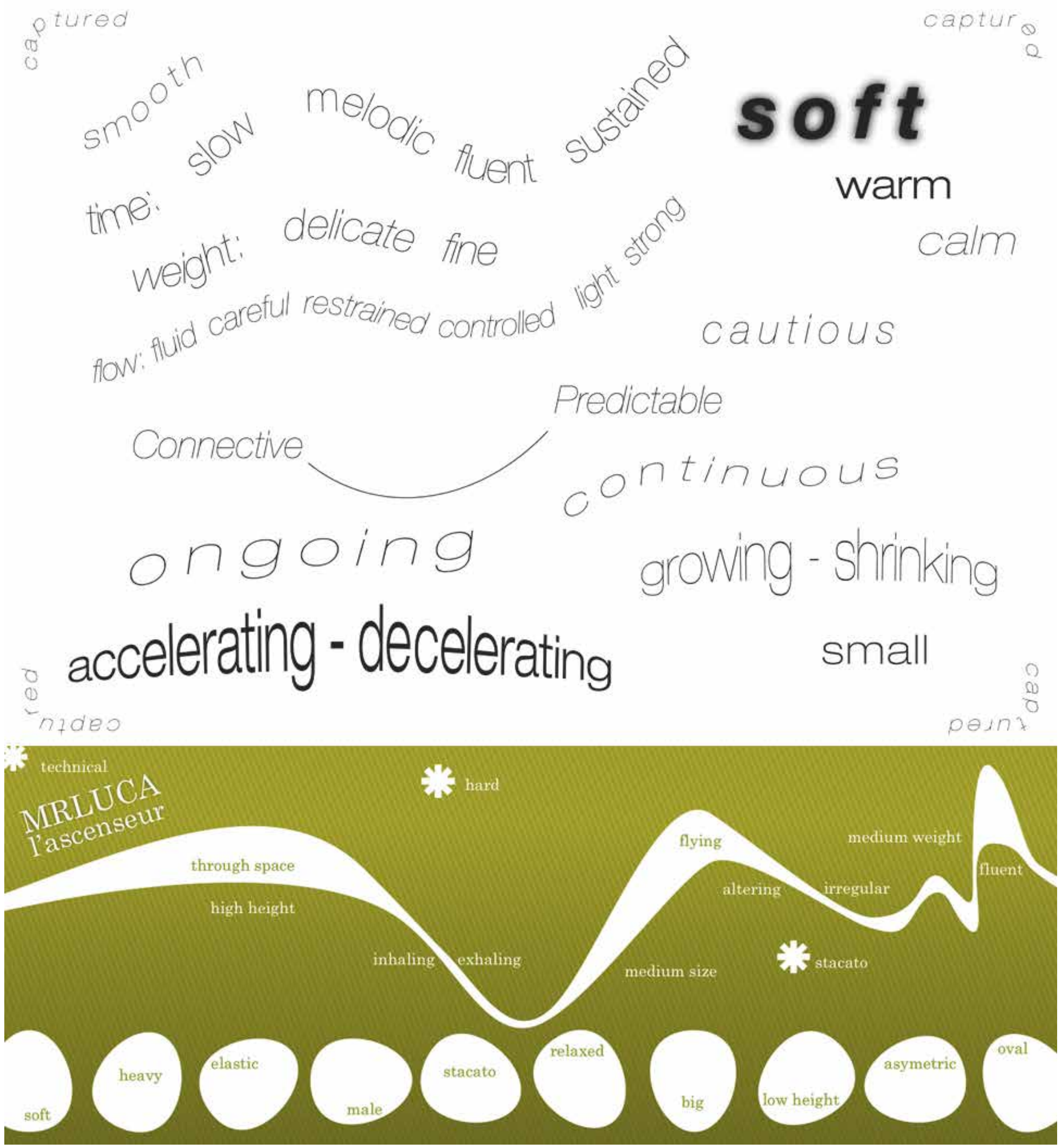



Figure 6.26: Interaction Maps: Students abstracted the movies into interaction maps with keywords and graphics.

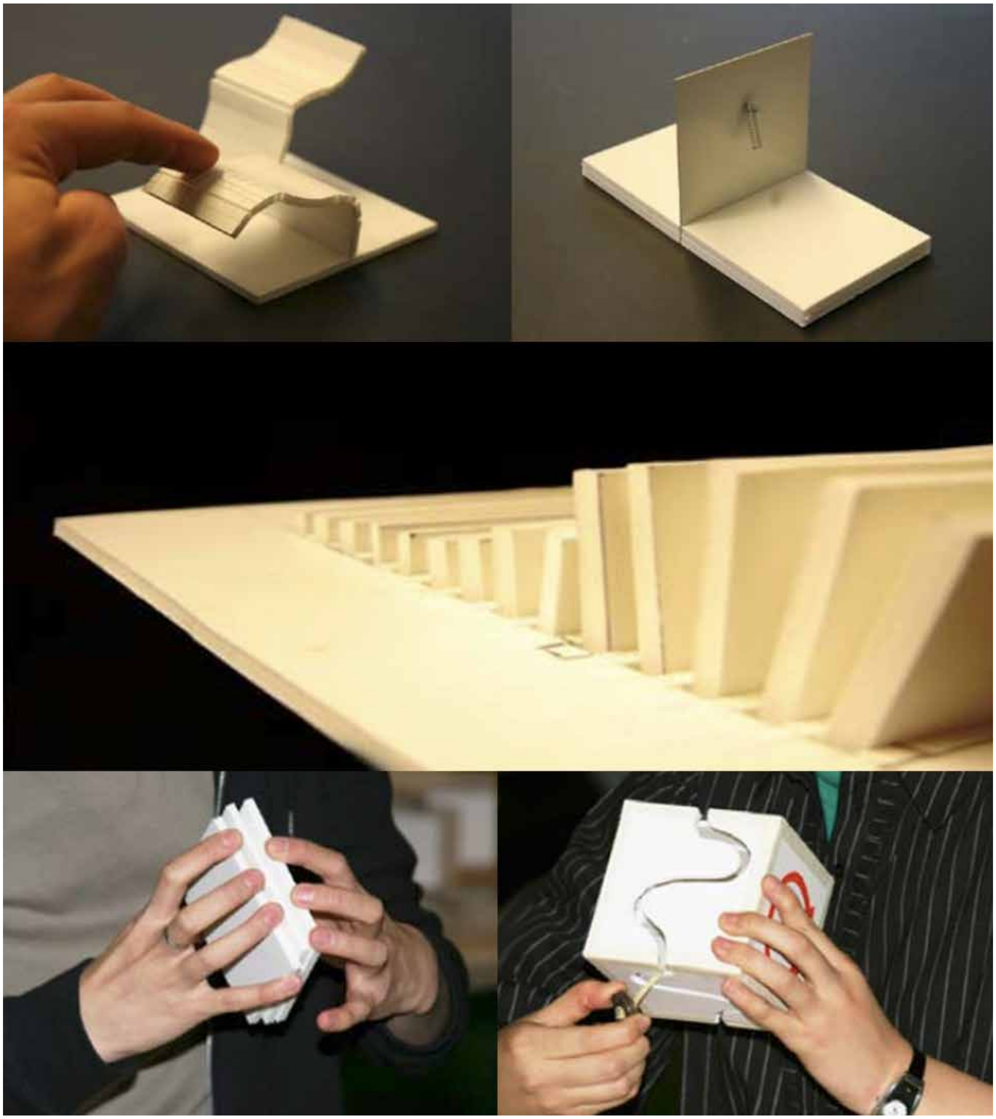


Figure 6.27: Interaction Mechanisms: Students translated the keywords into interaction possibilities. They built interaction mechanisms with white 'neutral' foam board squares in order to focus on interaction instead of appearance (sensing and abstracting).

Figure 6.28: Prototyping: Students built a prototype of a 'product' where the interaction had the same quality/experience potential as the music (cycles of doing and reflecting).

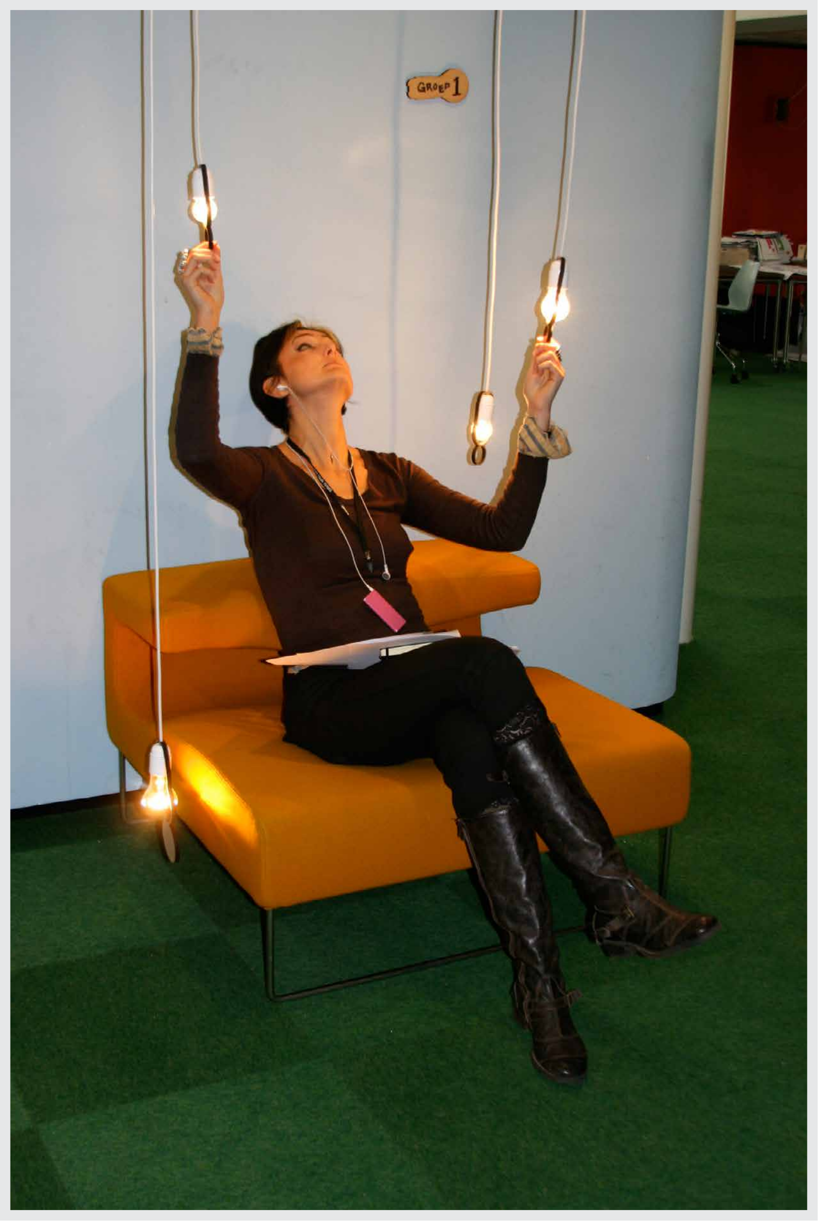



Figure 6.29: Testing: In a matching experiment, which the students set up, participants were asked to listen to the music pieces and then choose the matching prototypes (analysing and abstracting). We consider this last step to be important to interweave design teaching and research.

In the ongoing debate about design research and about how to generate knowledge, we argue for an intertwining of the sensorial/intuitive and the analytical/abstract approach (Hummels and Frens, 2008). What we see as quintessential, though, is the primacy of doing, as we have also explained in the principle 'primacy of action'. We believe there is already too much abstraction in the world that is not abstracted from reality, but from the abstract, e.g. from language, words and abstract concepts. Therefore, we want our students to be able to move confidently between these two, but always on the basis of the experiential, the real (Hummels et al, 2009).

#### 6.4.6.3 Related themes and research: 'first / third person perspective'

Points of view, including a first, second and third person perspective, is an essential element of phenomenology (see Part Three, Theoretical underpinning of industrial design). In her PhD work, Eva Deckers (to be published) uses a Research-through-Design approach to investigate if and how notions from the phenomenology of perception can inform design. Her approach is experimental phenomenology: the conditions for switching between first and third person are experimentally tested.

Deckers derives her inspiration from the work of Charles Lenay (2010). In his recent work, Lenay tries to understand the reasons for the pre-eminence of the tactile modality in descriptions of emotional exchanges. With the help of a minimalistic setup, he proposes a description of the essential conditions for 'contact' mediated by technical devices.

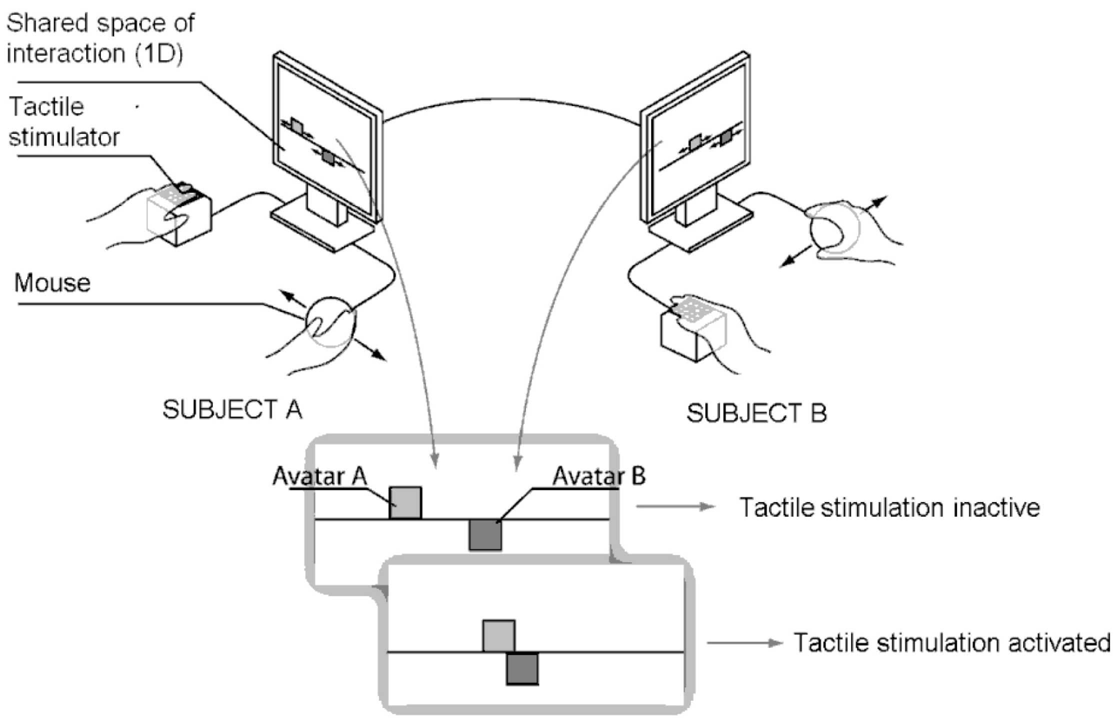



Figure 6.30: The one-dimensional space of perceptual interaction: with a computer mouse, each subject moves a receptor field on a line in a shared digital space. When the two receptor fields encounter each other, each user receives a tactile stimulus on his free hand (Lenay, 2010).



In order to comprehend the relationship between such contact and emotional values, Lenay understands emotion as being the product of a force that instigates movement. He then shows that the 'force' which is transmitted in touching contacts is based on the duality of the perceiving body and the body image of the self. The fact that the subject is ignorant of his own body-image (he cannot see his face) is revealed by the breaking of perceptual symmetry on the occasion of a touching encounter. Charles Lenay's results provide some guiding principles for the design of interfaces and structures of interaction that allow for mediated emotional contact.

These principles challenged Eva Deckers to incorporate them into the design of intelligent systems. In her research, she shows how perceptive behaviour is omni-present in our daily life and how it can be applied to design perceptive behaviour in objects. The goal is to provide knowledge and tools to design products or systems that are capable of perceptive activity and that can engage in reciprocal interplay with the perceived world, including multiple users and multiple products. She shows that this interplay between person and artefact positively influences the person's feeling of involvement in their common space. She designed a rug, PeR, that follows the caress of one's hand by leaving a light trail. Any behaviour can be given to the light in the rug. For example, it can flee from your touch, but also follow you closely (Deckers et al, 2012; or see the video at [http://dqi.id.tue.nl/per/PeR/Movie.html](http://dqi.id.tue.nl/per/PeR/Movie.html)).

FIGURE 6.31: PeR, Perception Rug, designed by Eva Deckers. PeR is a knotted carpet in which touch sensitivity and optic fibres are integrated. PeR has the ability to perceive and react to perceptive activity of people. PeR can adapt different perceptive behaviours.

Closely connected to PeR is a collection of sensual dynamics artifacts, developed by a group of Master students together with Pierre Levy, Eva Deckers, Michael Cruz (2012). These artifacts are able to sense a person and to behave upon it to invite for movements enhancing perceptive experience. Such an artifact is therefore the object of the experience, and at the same time the trigger for a greater sensory experience (see videos at http://dqi.id.tue.nl/sensual-dynamics/)

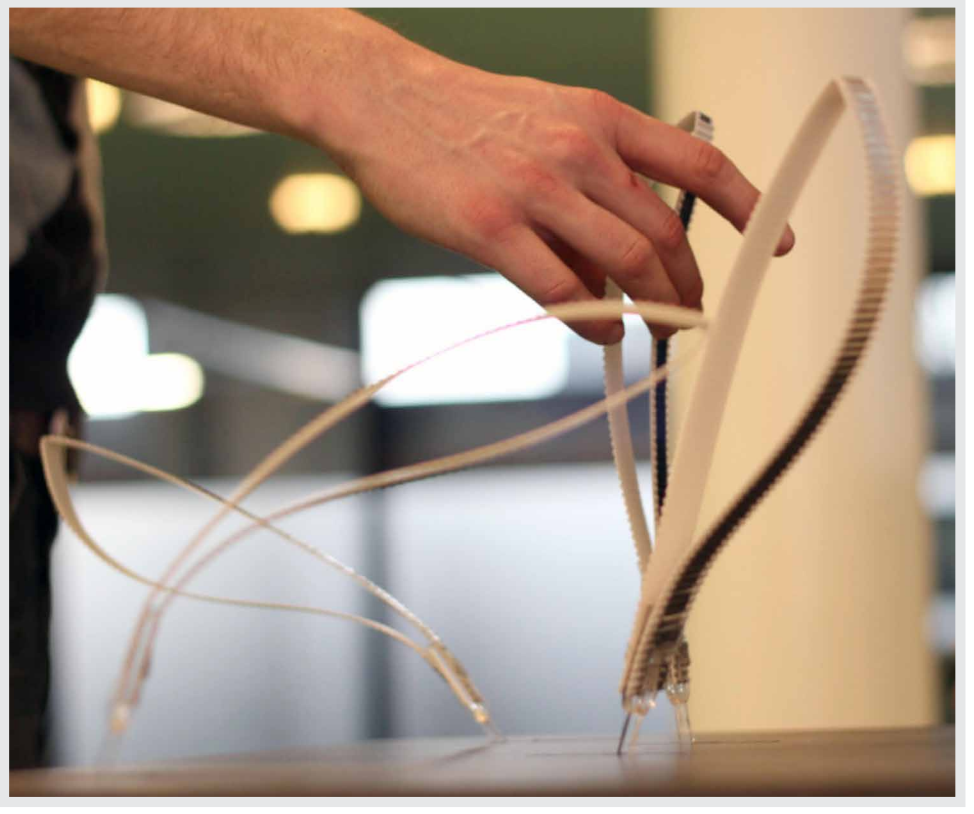


Figure 6.32: Be Touched, designed by Josje Wijnen, Jurrian Tjeenk Willink, Kim van Iersel and Sebastiaan Pijnappel. When one of the bodies is touched on the front it moves forward, in the direction of your hand, and you are touched back. At the backside the body is more ticklish. The body immediately moves away from your touch. Also, when one of the bodies is being touched the other bodies will start moving to draw your attention: they also want to be touched.

When looking at other related themes and research that address the first / third person perspective, one can see that many of the theories and frameworks described in this chapter apply, such as pragmatism, reflective practice, constructivist learning, reflective transformative design and also research-through-design. Bart Hengeveld's proposition that the designer-researcher is preferably one and the same person, since users interact with his design rationale, emphasises the importance of the first person perspective, a concept which is generally detested in classical research where the researcher is the objective observer. Being one and the same person enables the designer-researcher to easily switch from a first to a third person perspective and vice versa.

### 6.4.7 Principle 7. Creating opportunities for transformation through subtlety

Design can create new perspectives and approaches, and allows for transformation. Design is about creating opportunities instead of solving problems. Designers use aesthetic subtlety to catalyse meaning creation in interaction. To do so, designers use ambiguity, uncertainty, open-endedness, and resistance. They take risks and dare to fail.

#### 6.4.7.1 Example: The Other Brother

*Designed by: John Helmes (Master graduate TU/e)*
*Project team: Abigail Sellen (Microsoft Research Cambridge (UK) Lab) and Caroline Hummels (TU/e)*
*Time frame: 2009*
*Video:* http://www.youtube.com/watch?v=wVDXQuMJRac

'The Other Brother' is a semi-autonomous device that captures images and video of spontaneous moments in the course of everyday life to enable people to re-experience these moments in a playful way. It uses more serendipitous, lightweight ways to capture moments instead of conventional photo and video cameras, which require a person to take the initiative and to control the framing of the shot, leading to somewhat predictable results (Helmes et al, 2009).

The Other Brother is the result of an iterative process using the RTD process (Hummels and Frens, 2011), starting from initial (interactive) sketches, concepts and physical explorations over several prototypes and ending in a final design that was tested several times for a few weeks in a home environment.

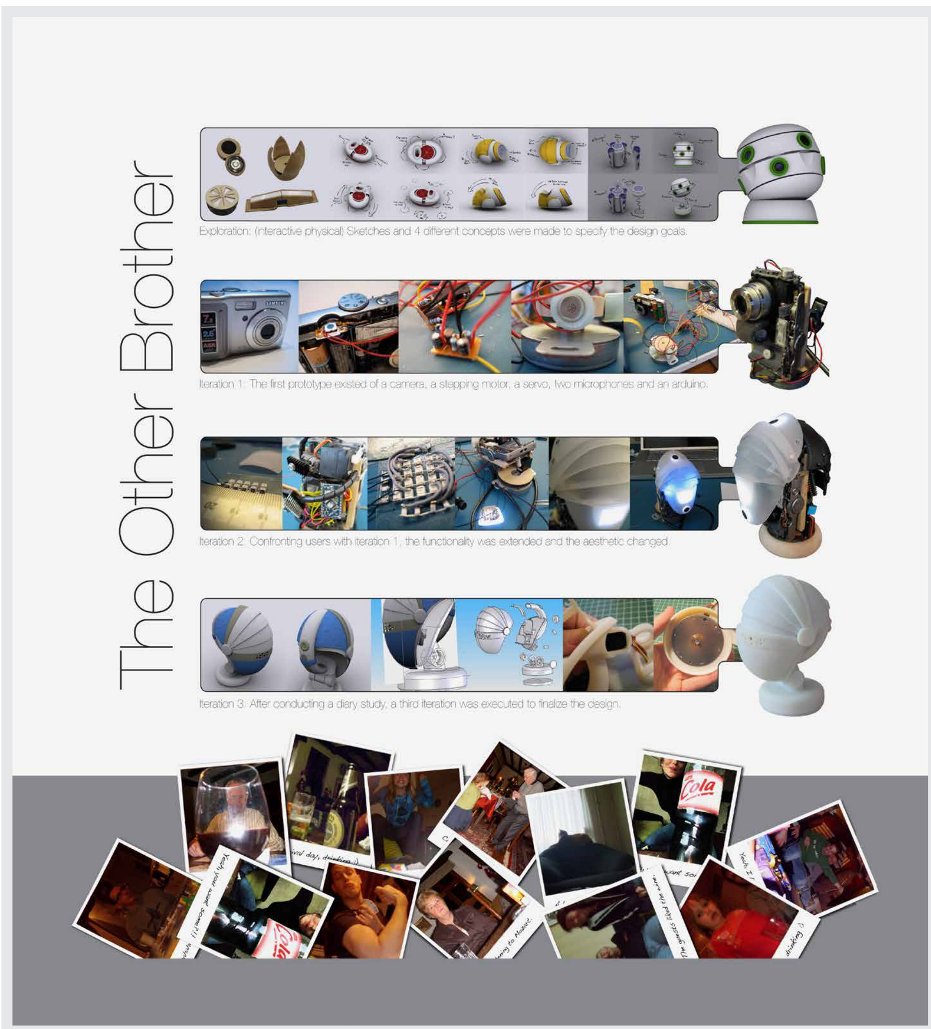



Figure 6.33: The development of the Other Brother through at least three experienceable working prototypes.

The project's goal was to design a disruptive innovative system in which disruptive refers to the absence of a well-established frame of reference for users, the market and society. Since disruptive innovations are open-ended and do not start with problem definition or finding, the system is all about creating opportunities for people to change their lives; opportunities through which a person can create meaning in a specific socio-cultural context. Since meaning is created in interaction between a person and an artefact, designers can 'merely' produce proposals for action and experience through their designed artefacts, while respecting people's perceptual-motor, social, emotional and cognitive skills.

Helmes tested the Other Brother in different families for one week in a row. The first test family existed of five people, a 50 year old father and mother, a 17 year old daughter and two sons, one aged 22 and the other 19. Initially, the father and mother of the family were quite sceptical about using the object since it was difficult for them to understand what exactly was being recorded and when. Since they really valued their privacy, their behaviour was not totally natural while the Other Brother was switched on. This is in strong contrast to the children, who did not experience any difficulty concerning privacy at all. Furthermore, the daughter seemed much more open to using the object in all kinds of different situations in comparison with her father. She also liked the fact that the Other Brother was on all the time and you simply did not pay any attention any more. This made the result very exciting since she did not know exactly what was captured. Another remarkable observation was the fact that she felt accompanied by the object while she spent an afternoon alone: "It was sociable, it made me feel not alone, I had something to talk to....I liked it staying with us for a whole week", she commented. In contrast to her father, she and her two brothers also intentionally triggered the Other Brother by making weird sounds or clapping their hands. As can been seen, the Other Brother was open enough for the children to explore new possibilities.

What is sometimes underestimated is the necessity of quality, subtlety and poetry of the design in interaction to catalyse meaning. Brewer et al (2007) state that the legibility/ambiguity tension concerns the extent to which a device is broadly understandable, but retains enough mystery both to be engaging and to allow users to project their own meanings onto it. We believe that design goes beyond pyramids, cubes and spheres, and beyond primary colours, which Brewer, Williams and Dourish used, and subsequently it will also be richer and even more effective.

Since products and systems are getting more intelligent, are taking more initiatives and seem at times to become 'part of the family' and thereby can (and probably will) transform society considerably, these developments require ethical discussion. Designers can be seen as catalysts for change by raising big societal questions as a way to transform society. In contrast to incremental changes, design can be used as inspiration for showing new ways, perspectives and approaches.

The goal of John Helmes' project was to envision and enable transformation. He deliberately used ambiguity, uncertainty and open-endedness in this design and design process. Throughout the several studies and workshops in the project, it became evident that the use of the Other Brother was emergent, i.e. its use was not *prescribed* by the designer — it simply *emerged*.

The project also surprised the client Microsoft Research Cambridge Lab: "By far the biggest surprise for us as a group was the realisation that the device itself appeared to have a life-like quality to it." Their responses clearly show that disruptive innovation, emergent behaviour and emergent interaction cannot be predicted or reasoned about beforehand. It is something that grows in interaction, while making, envisioning, testing, analysing, creating, etcetera in a real life context during the entire design process. Moreover, it requires a specific attitude from the designer using ambiguity, uncertainty, open-endedness, and resistance. The fact that Helmes took risks and dared to fail enabled this successful project.

### 6.4.7.2 Related themes and research

Transformation has become one of the buzzwords in design. In 2004, the British Design Council set up RED, a 'do tank' — as opposed to 'think tank' — that uses innovative design to tackle social and economic issues. They challenged accepted thinking and used creativity in close cooperation with users who were part of the design team to design new public services, systems and products that address social and economic problems. And they called their approach 'transformation design' (Burns et al, 2006).

Also in 2004, Matthijs van Dijk stated that the industry and, consequently, designers need to put more emphasis on true innovation based on a context-driven design strategy. Such a strategy evokes a paradigm shift by opening up and requiring a redefinition of the interaction between humans and products (Dijk van, 2004). Robert Fabricant, Vice President of Creative at Frog Design, sees a shift towards design with intent that has an immediate impact on user behaviour through direct social engagement (Fabricant, 2009). And Bruce Nussbaum, editor of the innovation and design coverage of *Business Week*, claimed that "'Innovation' is dead. Herald The Birth of 'Transformation' as The Key Concept for 2009". Nussbaum believes that transformation takes the best of design thinking and innovation and integrates them into a strategic guide for the unknowable and uncertain years ahead. He believes that we are already on the way to developing global networks working within ecosystems/ platforms (e.g. iTunes/iPod/iPhone, Nike Plus and Zipcar) that will make up our socio-economic and political worlds. Nussbaum considers transformation interesting because it approaches uncertainties with a methodology that creates options for new situations and sorts quickly through them for the best. It deals with a creative society in which we are all producers and consumers of value. Moreover, it can look at our systems — education, health-care, economic growth, transportation, defence, and political representation (Nussbaum, 2008).

### 6.4.7.2.1 OUT OF CONTROL

Designing for transformation requires specific design approaches based on open-endedness. It also requires enough self-confidence to design without having an idea of the solution space. It is about trusting one's intuition and senses (see principle 5), and as Sennett (2008) says, designing is about localising (making a matter concrete), questioning (reflecting on its quality) and opening up (expanding its sense). Designers should deliberately use ambiguity, uncertainty, open-endedness, complexity and resistance, and they should dare to fail. It is about being-out-of-control and being undisciplined, but still systematic during a creative, hands-on design process. This is essential for creating a complex unnatural (designed) world (Nelson, 1994). As we described earlier, development, both of products and of the designer himself, follows an equilibrium — disequilibrium — re-equilibrium pattern (Piaget, 1971). One goes from one stable state, equilibrium, to another, disequilibrium/chaos, and back to stability and order (re-equilibrium). The disequilibrium is the driving force of changing behaviour, transformation and development. Reflection and action are essential elements in regaining order because they can change ways of looking at the world and dealing with it. The re-equilibrium is a stable situation, but a design is never finished or right since the conditions, the situation and the user determine its status. So, designing is an activity that is highly iterative and dynamic.

Especially nowadays, when moving towards large living systems that are of a distributed, adaptable, evolvable, resilient and boundless nature, the number of variables and apparent connectivity become impossible to manage (Kelly, 1994). Moving towards such complexity implies that the challenges cannot be formulated exhaustively and that both challenges and solutions are not simply false or true; challenges are unique and there are multiple opportunities for solution spaces (Rittel, 1972). Consequently, designing complex systems can never be tackled through problem solving in a linear controlled process. Kelly (1994) states that the only way to develop or manage complex systems is by letting go of control and enabling the system to evolve without a central authority or imposed control mechanism. Although this last take might be too bold for many designers and even goes against the core of many design traditions, it stresses the need for an open process that supports evolving systems. According to Gilbert Cockton from Northumbria University (personal communication), this means that it cannot be based on an engineering approach (which starts from a problem), applied arts approach (which ignores the societal setting) or human-centred design approach (which is based on people's needs and wants).

#### 6.4.7.2.2 NEW APPROACHES IN A TRANSFORMATION ECONOMY

We believe that our complex societal challenges require solutions that support people in changing their behaviour on a personal and societal level. Since the complexity of designing such systems, products and services has increased rapidly during the last decade, the need for collaboration among all the cross-disciplinary stakeholders is becoming paramount. Reon Brand and Simona Rocchi (Brand and Rocchi 2011) see this kind of collaboration at the core of a new emerging economy, i.e. the transformation economy, which is based upon our current knowledge economy. Our current economy revolves around knowledge and information, which comes in different forms, e.g. shared information and reviews from users on Internet, sensor data from networked appliances, and knowledge from experts. However, our major societal challenges are leading to a growing discomfort and quest for balance and sustainability that cannot be created on an individual level only, as is the case in the knowledge paradigm. It requires behavioural change on a societal level too, where the collective is more important than the individual. These trends will move our society towards the transformation paradigm in the future, where stakeholders work together on local solutions for local issues that stem from greater global issues (Gardien et al, 2012).

Collaboration within the transformation economy requires engagement and empathy. It requires respect based on a horizontal collaboration in which all stakeholders are equal, but not identical, and each valuable in their own way. It asks people to put themselves - their point of view, their value system, their experience and their skills - into the shared design space. Such a transformation economy requires new ways of organising and working together, new business and financial models and new legal constructions. Through fostering continued relationships with local partners over time, organisations become respected players in these value networks (Hummels, 2012).

For example, within the context of the Dutch Creative Industry Scientific Programme (CRISP 2011), Oscar Tomico, Martijn ten Bhömer and Kristi Kuusk are developing a value co-creation platform that supports the development of innovative Product Service Systems (PSS) in the context of smart textile services. In order to face the complexity of PSS, this platform aims to facilitate close collaboration between small and medium enterprises from Dutch textile and technology industry, service partners, creative hubs and universities (Bhömerten, 2012).

Moreover, designers need new approaches, tools and techniques to face our societal challenges. Since meaning is created in interaction, it is impossible to predict whether the resulting outcome of designing for radical innovation will realise long-term societal change. Therefore, we have taken the design process into everyday life, involving a large group of stakeholders, including citizens in their everyday environment, thus realising valuable propositions together. To do this we have created Experiential Design Landscapes (EDL), infrastructures in neighbourhoods where all stakeholders work together, creating experienceable propositions with citizens, which evolve over time. These propositions, Experiential Probes (EP), are open, sensor-enhanced, networked products-service systems that enable citizens to develop new and emerging behaviour and, in parallel, enable detailed analysis of the emerging data patterns by researchers and designers as a source of inspiration for the development of future systems, products and services (Gent et al, 2011; Megens et al, 2012). This method is highly suitable for use within a transformation economy.

### 6.4.7.3 Future directions of industrial design

The field of industrial design is changing. That is the message when going to conferences such as TED, the World Design Forum (WDF), ICSID World Design Congress and CHI. According to Stefano Marzano, CEO of Philips Design, we are moving towards a new intellectual renaissance based on humanistic values. Designers are catalysts for change and raise large societal questions (Marzano, 2010). Consequently, the scope of design is changing. It is expanding towards all kind of systems: education, health-care, economic growth, transportation, defence, and political representation. Moreover, the role of designers is changing. Designers are dealing with a creative society in which we are all producers and consumers of value (Nussbaum, 2008).

As stated at the beginning of this chapter, industrial design is concerned with mass production in factories for the broadest market possible. However, given the developments in technology and society, *mass production* has extended to *mass customisation,* using new production techniques, and the broadest market possible has moved towards a diverse, global society that is no longer based on passive consumers; a society in which people are becoming accomplices in a never-ending design process of highly dynamic, adaptive and intelligent systems, products and services.

Figure 6.34: One example of the shift from mass production to mass customisation is a custom-made headphone made by Brian Garret (http://briangarret.com) for his Master graduation project (coached by Joep Frens, TU/e).



We see a desire for design-driven innovation; that is, we step away from incremental innovation in favour of disruptive innovation, in which disruptive refers to the absence of a well-established frame of reference for users or the market. Not only is the product as such new, but it also enables the creation of radical new meaning for the user, the market and society. The field is moving towards designing open and intelligent systems that evolve during use and have a high level of complexity due to their adaptive, context-dependent and highly dynamic character. Alongside this development, the role of the designer is changing. More and more we see open platforms and design projects in which a variety of people and experts create products. Finally, we see a changing attitude towards growth and development, towards self-directed learning and continuous development as a design professional.

Many question the appropriateness of the term 'industrial design' for this emerging field of expertise. However, since meaning is created in interaction, the concept of industrial design is also dynamic and changing. So, instead of changing its name, we believe we should explore the potential of industrial design, as we have tried to do over the last twenty years and have documented in this chapter.

## 6.5 ACKNOWLEDGEMENTS

The work described in this chapter is based on well over twenty-five years of design research and education, and it has been - and is still - a group effort. Therefore, we would like to thank all our friends from the Designing Quality in Interaction group at Eindhoven University of Technology, as well as our national and international friends, colleagues, students and alumni with whom we collaborated closely to develop these principles of industrial design.

## 6.6 REFERENCES

*(1990): Technology and the Lifeworld : From Garden to Earth.*

Achterhuis, *Hans (1997): Van Stoommachine tot cyborg; denken over techniek in de nieuwe wereld. Ambo*

Anderson, *Lorin W. and* Krathwohl, *David R. (2000): A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives, Abridged Edition. Pearson*

Archer, *Bruce (1995): The Nature of Research. In* Journal of Co-Design,

Baker, *Robin (1993): Designing the Future: The Computer in Architecture and Design. Thames and Hudson*

*Bartneck, Christoph and Rauterberg, Matthias (2007): HCI reality -- an 'Unreal Tournament'?. In International Journal of Human-Computer Studies, 65 (8) pp. 737-743*

*Bhöme, Martijn ten, Tomico, Oscar, Kleinsmann, Maaike, Kuusk, Kristi and Wensveen, Stephan (2012): Designing Smart Textile Services through value networks, team mental models and shared ownership. In: Service design and innovation conference ServDes. 2012 8-10 February, 2012, Espoo, Finland. pp. 1-10*

*Bijker, Wiebe E. and Law, John (eds.) (1994): Shaping Technology / Building Society: Studies in Sociotechnical Change (Inside Technology). The MIT Press*

*Birenbaum, Menucha (2003): New Insights Into Learning and Teaching and Their Implications for Assessment. In: Segers, Mien, Dochy, F. and Cascallar, E. (eds.). "Optimising New Modes of Assessment: In Search of Qualities and Standards (Innovation and Change in Professional Education)". Springer*

*Bos, Maarten W., Baaren, Rick B. van, Nordgren, Loran F. and Dijksterhuis, Ap (2006): Making choices without deliberation. In Science, 312 (5779) pp. 1472-1472*

*Brand, Reon and Rocchi, Simona (2011). Rethinking value in a changing landscape; a model for strategic reflection and business transformation. Retrieved 9 November 2012 from design.philips.com: http://www.design.philips.com/philips/shared/assets/design_assets/pdf/nvbD/april2011/paradigms.pdf*

*Brewer, Johanna, Williams, Amanda and Dourish, Paul (2007): A handle on what's going on: combining tangible interfaces and ambient displays for collaborative groups. In: Ullmer, Brygg and Schmidt, Albrecht (eds.) Proceedings of the 1st International Conference on Tangible and Embedded Interaction 2007, Baton Rouge, Louisiana, USA, February 15-17, 2007 2007. pp. 3-10*

*Burdek, Bernhard E. (2005): Design: The History, Theory and Practice of Product Design. Birkhäuser Architecture*

*Burns, Colin, Cottam, Hilary, Vanstone, Chris and Winhall, Jennie (2006). Red paper 02; Transformation Design. Retrieved 9 November 2012 from Designcouncil.info: http://www.designcouncil.info/mt/RED/transformationdesign/TransformationDesignFinalDraft.pdf*

*Buur, Jacob, Jensen, Mads Vedel and Djajadiningrat, Tom (2004): Hands-only scenarios and video action walls: novel methods for tangible user interaction design. In: Proceedings of DIS04: Designing Interactive Systems: Processes, Practices, Methods, & Techniques 2004. pp. 185-192*

*Coker, William and Shook, Jan (2006): Increasing the appeal of AAC technologies using VSD's in preschool language intervention. In: Proceedings of the 22nd Annual International Technology and Persons with Disabilities Conference 19-24 March, 2006, Los Angeles, USA.*

*Coyne, Richard and Snodgrass, Adrian (1993): Cooperation and individualism in design. In Environment and planning B, 20 (2) p. 163–174*

*Cross, Nigel (2011): Design Thinking: Understanding How Designers Think and Work. Berg Publishers*

*Cross, Nigel, Christiaans, Henri and Dorst, Kees (eds.) (1996): Analysing Design Activity. London, Wiley*

*Deckers, Eva, Levy, Pierre, Wensveen, Stephan, Ahn, Rene and Overbeeke, Kees (2012): Designing for perceptual crossing: applying and evaluating design notions. In International Journal of Design, 6 (3)*

*Dewey, John (1934): Art as Experience. Perigee Trade*

*Dewey, John (1997): Experience And Education. Free Press*

*Dijk, Jelle van (2012). Sharing traces, making sense: How interactive technology shapes embodied cognitive group practices – and vice versa. University of Technology Eindhoven*

*Dijksterhuis, Ap, Bos, Maarten W., Nordgren, Loran F and Baaren, Rick B. van (2006b): On making the right choice: the deliberation-without-attention effect. In Science, 311 (5763) pp. 1005-1007*

*Dijksterhuis, Ap, Bos, Maarten W., Nordgren, Loran F. and Baaren, Rick B. van (2006c): Complex choices better made unconsciously?. In Science, 313 pp. 760-761*

*Dijksterhuis, Ap, Bos, Maarten W., Nordgren, Loran F. and Baaren, Rick B. van (2006a): Making choices without deliberation. In Science, 312 p. 1472*

*Djajadiningrat, J. P., Overbeeke, Kees and Wensveen, Stephan (2000): Augmenting fun and beauty: a pamphlet. In: Designing Augmented Reality Environments 2000 2000. pp. 131-134*

*Djajadiningrat, Tom, Wensveen, Stephan, Frens, Joep and Overbeeke, Kees (2004): Tangible Products: redressing the balance between appearance and action. In Personal and Ubiquitous Computing, 8 (5) pp. 294-309*

*Djajadiningrat, Tom, Matthews, Ben and Stienstra, Marcelle A. (2007): Easy doesn't do it: skill and expression in tangible aesthetics. In Personal and Ubiquitous Computing, 11 (8) pp. 657-676*

Doll, William E. (1986): Prigogine: A New Sense of Order, A New Curriculum. In Theory into Practice, 25 (1) pp. 10-16

Dourish, Paul (2001): Where the Action Is: The Foundations of Embodied Interaction. MIT Press

Dunne, Anthony and Raby, Fiona (2001): Design Noir: The Secret Life of Electronic Objects. Birkhäuser Science

Fabricant, Robert (2009). Design with intent. How designers can influence behavior. Retrieved 11 November 2012 from Designmind.frogdesign.com: http://designmind.frogdesign.com/articles/power/design-with-intent.html

Fabricant, Robert (2009a): Project Masiluleke. In Interactions, 16 (6) pp. 6-12

Fitzmaurice, George W., Ishii, Hiroshi and Buxton, Bill (1995): Bricks: Laying the Foundations for Graspable User Interfaces. In: Katz, Irvin R., Mack, Robert L., Marks, Linn, Rosson, Mary Beth and Nielsen, Jakob (eds.) Proceedings of the ACM CHI 95 Human Factors in Computing Systems Conference May 7-11, 1995, Denver, Colorado. pp. 442-449

Forlizzi, Jodi and Battarbee, Katja (2004): Understanding experience in interactive systems. In: Proceedings of DIS04: Designing Interactive Systems: Processes, Practices, Methods, & Techniques 2004. pp. 261-268

Forty, Adrian (1986): Objects of Desire: Design and Society Since 1750. Thames and Hudson

Frens, Johannes W. (2006). Designing for rich interaction: Integrating form, interaction and function. Industrial Design-University of Technology Eindhoven http://alexandria.tue.nl/extra2/200610381.pdf

Gardien, Paul, Djajadiningrat, Tom, Hummels, Caroline and Brombacher, Aarnout C. (2012): Innovation paradigms: how design needs to evolve to deliver value. In International Journal of Design, 6 (3)

Gaver, William W. (1991): Technology Affordances. In: Robertson, Scott P., Olson, Gary M. and Olson, Judith S. (eds.) Proceedings of the ACM CHI 91 Human Factors in Computing Systems Conference April 28 - June 5, 1991, New Orleans, Louisiana. pp. 79-84

Gent, Sabine van, Hummels, Caroline, J.P.G.Megens, Carl, Peeters, Michel, Lu, Yuan and Brombacher, Aarnout C. (2011): Experiential Design Landscapes as a design tool for market research of disruptive intelligent systems. In: Best, Kathryn, Bruce, Margaret, Buijs, Jan, Cooper, Rachel and Gemser, Gerda (eds.) Proceedings of the 1st Cambridge Academic Design Management Conference September 7-8, 2011, Cambridge, United Kingdom .

*Gibson, James J. (1979): The Ecological Approach to Visual Perception. New Jersey, USA, Lawrence Erlbaum Associates*

*Heidegger, Martin (1962): Being and Time. English translation 1962.*

*Heidegger, Martin (1927): Being and Time. Harper Perennial Modern Classics*

*Helmes, John, Hummels, Caroline and Sellen, Abigail (2009): The other brother: re-experiencing spontaneous moments from domestic life. In: Villar, Nicolas, Izadi, Shahram, Fraser, Mike and Benford, Steve (eds.) TEI 2009 - Proceedings of the 3rd International Conference on Tangible and Embedded Interaction February 16-18, 2009, Cambridge, UK. pp. 233-240*

*Hengeveld, Bart J. (2011). Designing LinguaBytes : a tangible language learning system for non- or hardly speaking toddlers. University of Technology Eindhoven http://alexandria.tue.nl/extra2/715535.pdf*

*Heskett, John (1980): Industrial Design (World of Art Series). W W Norton and Co Inc*

*Hine, Thomas (1986): Populuxe. Overlook TP*

*Horn, Richard (1985): Memphis: Objects, Furniture and Patterns. Simon and Schuster (Paper)*

*Hummels, Caroline (2000b). Evoke experiences through engagement. Technical University Delft http://dqi.id.tue.nl/docs/Hummels2000.pdf*

*Hummels, Caroline (2000a). Gestural design tools: prototypes, experiments and scenarios. Delft University of Technology http://id-dock.com/pages/overig/caro/publications/thesis/00Humthesis.pdf*

*Hummels, Caroline (2012). Matter of transformation: sculpting a valuable tomorrow. Retrieved 11 November 2012 from University of Technology Eindhoven: http://dqi.id.tue.nl/docs/Inaugural_Lecture-Caroline_Hummels-2012.pdf*

*Hummels, Caroline (2007): Searching for Salient Aspects of Resonant Interaction. In Knowledge Technology Policy, 20 (1) pp. 19-29*

*Hummels, Caroline and Frans, Joep W. (2008): Designing for the unknown: a design process for the future generation of highly interactive systems and products,. In: Proceedings of the 10th International Conference on Engineering and Product Design Education 2008. pp. 204-209*

*Hummels, Caroline and Frens, Joep (2009): The reflective transformative design process. In: Proceedings of ACM CHI 2009 Conference on Human Factors in Computing Systems 2009. pp. 2655-2658*

*Hummels, Caroline and Frens, Joep W. (2008): Designing for the unknown: A design process for the future generation of highly interactive systems and products. In Reflective Practice, 4 (0) pp. 1-6*

*Hummels, Caroline and Frens, Joep (2011): Designing Disruptive Innovative Systems, Products and Services: RTD Process. In: Coelho, Denis (ed.). "Industrial Design-New Frontiers". Rijeka, Croatia: InTechpp. 145-172*

*Hummels, Caroline and Overbeeke, Kees (2010): Special Issue Editorial: Aesthetics of Interaction. In International Journal of Design, 4 (2) pp. 1-2*

*Hummels, Caroline and Vinke, Diana (2009): Eindhoven designs; volume two: Developing the competence of designing intelligent systems. Eindhoven, The Netherlands, University of Technology Eindhoven*

*Hummels, Caroline, Overbeeke, Kees and Klooster, Sietske (2007): Move to get moved: a search for methods, tools and knowledge to design for expressive and rich movement-based interaction. In Personal and Ubiquitous Computing, 11 (8) pp. 677-690*

*Hummels, Caroline, Restrepo, Michael Cruz and Overbeeke, Kees (2009): Designing the melody of interaction through movies, maps, mechanisms, prototypes and presentations. In: Proceedings of ACM CHI 2009 Conference on Human Factors in Computing Systems 2009. pp. 2647-2650*

*Ihde, Don (1990): Technology and the Lifeworld: From Garden to Earth (Indiana Series in the Philosophy of Technology). Indiana University Press*

*Ishii, Hiroshi and Ullmer, Brygg (1997): Tangible Bits: Towards Seamless Interfaces between People, Bits and Atoms. In: Pemberton, Steven (ed.) Proceedings of the ACM CHI 97 Human Factors in Computing Systems Conference March 22-27, 1997, Atlanta, Georgia. pp. 234-241*

*ISO standard 9241, Ergonomic requirements for office work with visual display terminals.*

*Kelly, Kevin (1994): Out of Control: The New Biology of Machines, Social Systems, & the Economic World. Basic Books*

*Kint, Johanna (2001): Expo 58 als belichaming van het humanistisch modernisme. 010*

*Kint, Johanna M. L., Overbeeke, Kees J. and Ross, Philip (2010): D'ou venons nous? Que sommes nous? Ou allons nous?. In: Aoussat, Ameziane, Shiizuka, Hisao and Chen, Kuohsiang (eds.) International Conference on Kansei Engineering and Emotion Research March 2-4, 2010, Paris, France.*

*Koskinen, Ilpo, Zimmerman, John, Binder, Thomas, Redstrom, Johan and Wensveen, Stephan (2011): Design Research Through Practice: From the Lab, Field, and Showroom. Morgan Kaufmann*

*Kyffin, Steven, Feijs, Loe and Djajadiningrat, Tom (2005): Exploring expression of form, action, and interaction. In: Sloane, Andy (ed.). "Home-Oriented Informatics and Telematics: Proceedings of the IFIP WG 9.3 HOIT2005 Conference (IFIP Advances in Information and Communication Technology)". Springer*

*Latour, Bruno (1994): Where are the missing masses? The sociology of a few mundane artifacts. In: Bijker, Wiebe E. and Law, John (eds.). "Shaping Technology / Building Society: Studies in Sociotechnical Change (Inside Technology)". The MIT Press*

*Lawson, Bryan (2005): How Designers Think, Fourth Edition: The Design Process Demystified. Architectural Press*

*Lawson, Bryan and Dorst, Kees (2009): Design Expertise. Taylor and Francis*

*Lenay, Charles (2010): It's So Touching: Emotional Value in Distal Contact. In International Journal of Design, 4 (2) pp. 15-25*

*Locher, Paul, Overbeeke, Kees and Wensveen, Stephan A. G. (2010): Aesthetic Interaction: A Framework. In Design Issues, 26 (2) pp. 70-79*

*Maeda, John (ed.) (1999): Design by Numbers. The MIT Press*

*Marti, Patrizia and Overbeeke, Kees (C. J.) (2011): Designing complexity in context: Light through Culture. In: Proceedings of the 9th ACM SIGCHI Italian Chapter International Conference on Computer-Human Interaction Facing Complexity 2011. pp. 65-70*

*Marzano, Stefano (2011). The next design. Retrieved 12 November 2012 from worlddesignforum.nl: http://www.worlddesignforum.nl/?p=15*

*Megens, Carl J. P. G., Peeters, Michel M. R., Hummels, Caroline and Brombacher, Aarnout C. (2012): Designing experiential probes for the wild. In ACM Transactions on Computer-Human Interaction,*

*Merleau-Ponty, Maurice (1958): Phenomenology of Perception. London, England, Routledge*

*Merleau-Ponty, Maurice (1962): Phenomenology of Perception: An Introduction. Routledge*

*Michaels, Claire F. and Carello, Claudia (1981): Direct Perception. Prentice Hall*

*Nelson, Harold G. (1994): The Necessity of Being "Un-Disciplined" and "Out-of-Control": Design Action and Systems Thinking. In Performance Improvement Quarterly, 7 (3) pp. 22-29*

*Norman, Donald A. (1988): The Psychology of Everyday Things. New York, Basic Books*

*Nussbaum, Bruce (2008). "Innovation" is Dead. Herald The Birth of "Transformation" as The Key Concept for 2009. Retrieved 1 August 2011 from BusinessWeek: http://www.businessweek.com/innovate/NussbaumOnDesign/archives/2008/12/innovation_is_d.html*

*Nussbaum, Bruce (2008). "Innovation" is dead. Herald the birth of "transformation" as the key concept for 2009. Retrieved 13 November 2012 from Businessweek.com: http://www.businessweek.com/stories/2008-12-30/innovation-is-dead-dot-herald-the-birth-of-transformation-as-the-key-concept-for-2009-dot*

*Overbeeke, Kees (2007). The aesthetics of the impossible. Inaugural lecture, October 26, 2007. Eindhoven University of Technology http://dqi.id.tue.nl/docs/Inaugural_Lecture-Kees_Overbeeke-2007.pdf*

*Overbeeke, Kees and Wensveen, Stephan (2003): From perception to experience, from affordances to irresistibles. In: DPPI 2003 - Proceedings of the 2003 International Conference on Designing Pleasurable Products and Interfaces June 23-26, 2003, Pittsburgh, PA, USA. pp. 92-97*

*Overbeeke, Kees, Djajadiningrat, Tom, Hummels, Caroline and Wensveen, Stephan (2002): Beauty in usability: Forget about ease of use!. In: Green, William S. and Jordan, Patrick W. (eds.). "Pleasure with Products: Beyond Usability". CRC Press*

*Overbeeke, Kees, Wensveen, Stephan A. G. and Hummels, Caroline (2006): Design research: generating knowledge through doing. In: Michel, Ralf and Hirt, Lysianne L. (eds.) Drawing new territories; best of design research, 3rd symposium of design research 17-18 November, 2006, Geneva, Switzerland . pp. 51-69*

*Petersen, Marianne Graves, Iversen, Ole Sejer, Krogh, Peter Gall and Ludvigsen, Martin (2004): Aesthetic interaction: a pragmatist's aesthetics of interactive systems. In: Proceedings of DIS04: Designing Interactive Systems: Processes, Practices, Methods, & Techniques 2004. pp. 269-276*

*Philips, Philips (1996): Vision of the future. Philips*

*Piaget, Jean (1971): Biology and Knowledge. University of Chicago Press*

*Rittel, Horst W. J. (1972): On the planning crisis: Systems analysis of the 'first and second generations'. In Bedriftsøkonomen, 8 pp. 390-396*

*Ross, Philip (2008). Ethics and aesthetics in intelligent product and system design. University of Technology Eindhoven http://www.idemployee.id.tue.nl/p.r.ross/thesis/Ethics%20and%20aesthetics%20site/Home_files/Ross-thesis.pdf*

*Sanders, John T. (1997): An ontology of affordances. In Ecological psychology, 9 (1) pp. 97-112*

*Schwartz, Shalom H. (1992): Universals in the content and structure of values: Theory and empirical tests in 20 countries. In: Zanna, Mark (ed.). "Advances in experimental social psychology". New York, USA: Academic Presspp. 1-65*

*Schwartz, Shalom H. (2004): Basic human values: Their content and structure across countries. In: Porto, Juliana B. and Tamayo, Álvaro (eds.). "Valores de trabalho [Values and work]". Universidad de Brasil*

*Schön, Donald A. (1983): The Reflective Practitioner: How Professionals Think In Action. Basic Books*

*Sennett, Richard (2008): The Craftsman. Yale University Press*

*Shaer, Orit and Hornecker, Eva (2010): Tangible User Interfaces: Past, Present, and Future Directions. In Foundations and Trends in Human-Computer Interaction, 3 (1) pp. 1-137*

*Shusterman, Richard (2008): Body Consciousness: A Philosophy of Mindfulness and Somaesthetics. Cambridge University Press*

*Shusterman, Richard (2000): Pragmatist Aesthetics: Living Beauty, Rethinking Art. Rowman and Littlefield Publishers, Inc*

*Smets, Gerda and Overbeeke, Kees (1994): Industrial design engineering and the theory of direct perception. In Design Studies, 15 (2) pp. 175-184*

*Snow, Catherine E. and Ferguson, Charles A. (eds.) (1977): Talking to Children: Language Input and Acquisition. Cambridge University Press*

*Sparke, Penny (1996): An Introduction to Design and Culture: 1900 to the Present. Routledge*

*Sparke, Penny (1986): An Introduction to Design and Culture: 1900 to the Present. Allen and Unwin*

*Stappers, Pieter J. (2007): Doing design as a part of doing research. In: Michel, Ralf (ed.). "Design research now: Essays and selected projects". Basel, Switzerland: Birkhäuserpp. 81-91*

*Stienstra, Jelle T., Bruns, Miguel, Wensveen, Stephan A. G. and Kuenen, Stoffel D. (2012): How to design for transformation of behavior through interactive materiality. In: Proceedings of the 7th Nordic Conference on Human-Computer Interaction NordiCHI 2012 October 14-17, 2012, Copenhagen, Denmark.*

*Stienstra, Jelle, Overbeeke, Kees and Wensveen, Stephan (2011): There is More in a Single Touch: Mapping the Continuous to the Discrete. In: Overbeeke, Kees (ed.) CHItaly Proceedings of the 9th ACM SIGCHI Italian Chapter International Conference on Computer-Human Interaction Facing Complexity 2011. pp. 27-32*

*Stolterman, Erik (2008): The nature of design practice and implications for interaction design research. In International Journal of Design, 2 (1) pp. 55-65*

*Title: Principes Author: Marten van Dijk Publication: Inaugural speech Year: 2004 Publisher: Delft University of Technology*

*Tomasello, Michael (2003): Constructing a Language: A Usage-Based Theory of Language Acquisition. Harvard University Press*

*Trotto, Ambra, Hummels, Caroline and Restrepo, Michael Cruz (2011): Towards design-driven innovation: designing for points of view using intuition through skills. In: Proceedings of Designing Pleasurable Products and Interfaces 2011, Milano. pp. 3-10*

*Verbeek, Peter P. (2006): Materializing morality – Design ethics and technological mediation. In Science, Technology and Human Values, 31 (3) pp. 361-380*

*Wakkary, Ron (2005): Framing complexity, design and experience: a reflective analysis. In Digital Creativity, 16 (2) pp. 65-78*

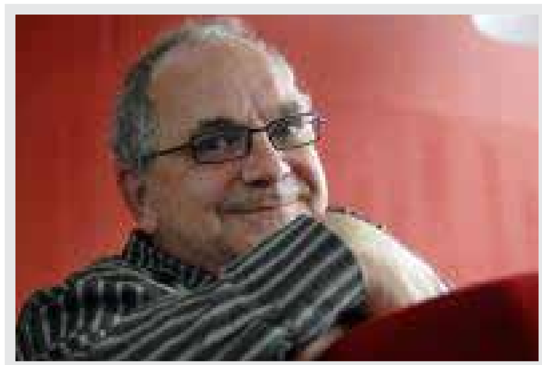
*Weiser, Mark (1991): The Computer for the 21st Century. In Scientific American, 265 (3) pp. 94-104*

*Wensveen, S. A. G., Djajadiningrat, J. P. and Overbeeke, Kees (2004): Interaction frogger: a design framework to couple action and function through feedback and feedforward. In: Proceedings of DIS04: Designing Interactive Systems: Processes, Practices, Methods, & Techniques 2004. pp. 177-184*

*Zimmerman, John, Forlizzi, Jodi and Evenson, Shelley (2007): Research through design as a method for interaction design research in HCI. In: Proceedings of ACM CHI 2007 Conference on Human Factors in Computing Systems 2007. pp. 493-502*

## About the authors

### Kees Overbeeke



© *Kees Overbeeke*

Has also published under the name of:
"C. J. Overbeeke" and "Kees C. J. Overbeeke"

Personal Homepage:
http://dqi.id.tue.nl/kees/

Prof. Dr. Kees Overbeeke (1952-2011) was full professor at Eindhoven University of Technology (TU/e) for Intelligent Products and System Design in the Department of Industrial Design in May, 2006. Kees Overbeeke studied psychology at the Katholieke Universititeit Leuven (1974). After working there he moved to the Faculty of Industrial Design Engineering at Delft University of Technology where he earned his Ph.D (1988) in spatial perception on flat screens. He headed the Form Theory group as an Associate Professor until his move to the Department of Industrial Design of TU/e in 2002. During the academic year 2005-2006 he was invited as the Nierenberg Chair of Design at Carnegie Mellon's School of Design in Pittsburgh. At TU/e he headed the Designing Quality in Interaction group until September 2011. It is with great sadness that we announce the loss of Kees Overbeeke, on October 8th, 2011 at the age of 59.

## Caroline Hummels

© *Caroline Hummels*

Has also published under the name of:
"Caroline C. M. Hummels"

Prof. dr. ir. Caroline Hummels is full professor of Design Theory at the department of Industrial Design, TU/e, as well as head of the Designing Quality in Interaction group. She studied Industrial Design Engineering at the Delft University of Technology (cum laude) and also obtained her PhD (cum laude) there. Her current activities concentrate on designing for personal, social and societal transformation, with a special focus on health and well-being. She designs for transformation based on the aesthetics of interaction with open, disruptive innovative systems within a societal context. Moreover, she is a designer-researcher with extensive experience in (interaction) design, education and research-trough design. She has developed various (research) prototypes and installations next to design methods, techniques and processes. She is a member of the steering committee of the Tangible Embedded, and Embodied Interaction Conference (TEI) as well as editorial board member of the International Journal of Design. Moreover, she has been a member of a variety of international program committees and national Think Tanks since 2001, and has given a large number of keynote speeches, invited lectures and workshops at conferences, international universities and for industry world wide.

## Your notes and thoughts on chapter 6

Record your notes and thoughts on this chapter. If you want to share these thoughts with others online, go to the bottom of the page at:
http://www.interaction-design.org/encyclopedia/industrial_design.html

**NOTES**

# Chapter 7

# Bifocal Display

*by Robert Spence and Mark Apperley.*

The Bifocal Display is an information presentation technique which allows a large data space to be viewed as a whole, while simultaneously a portion is seen in full detail. The detail is seen in the context of the overview, with continuity across the boundaries, rather than existing in a disjoint window (see Figure 7.1).

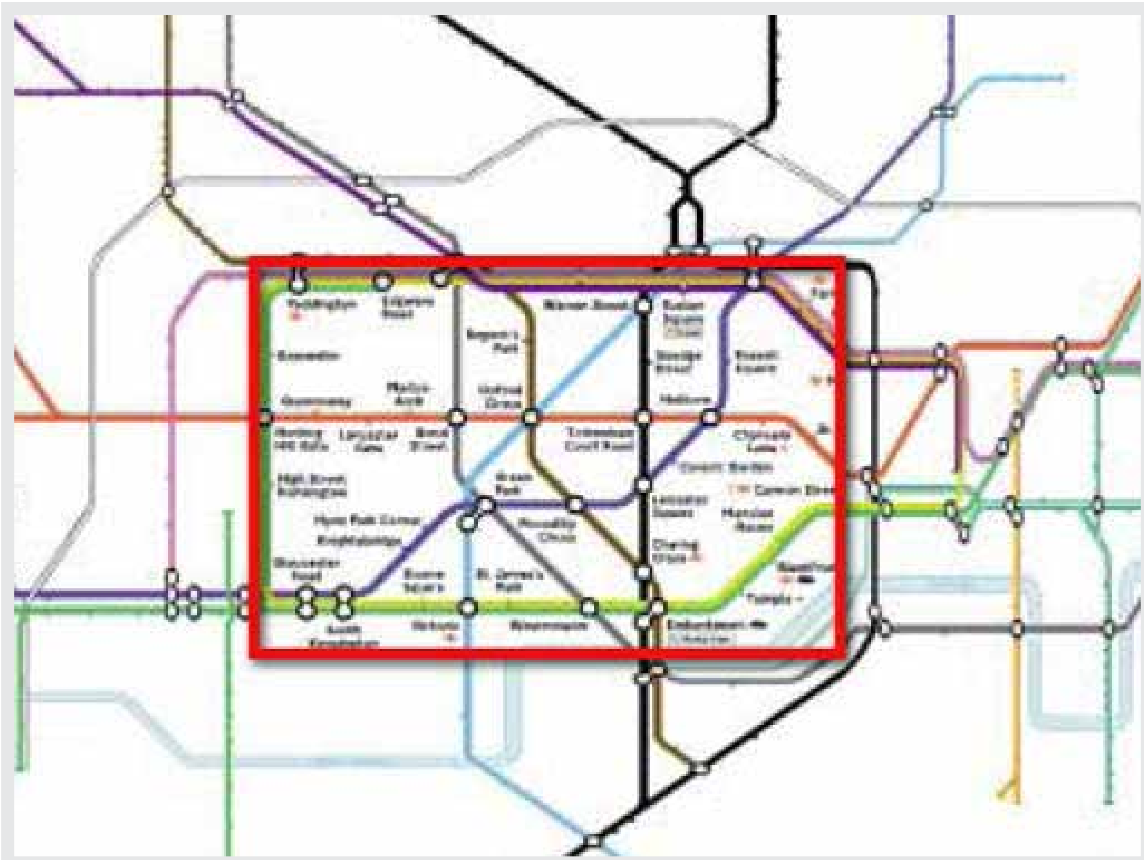
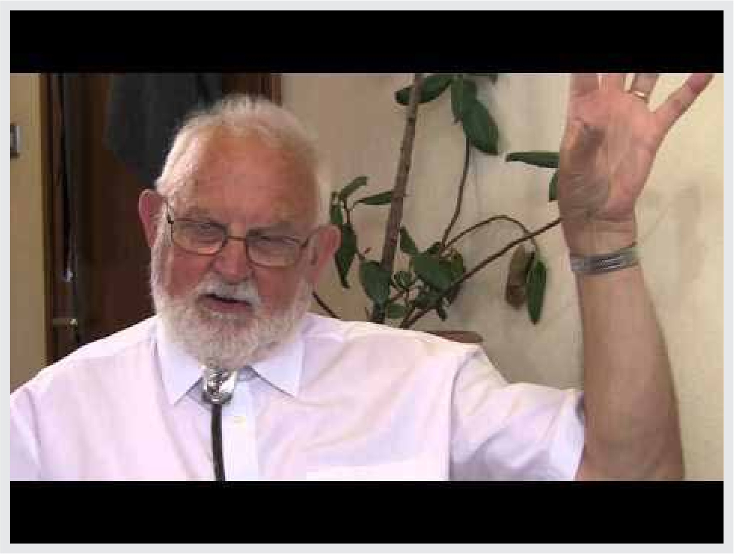


FIGURE 7.1: A bifocal representation of the London Underground map, showing the central area in full detail, while retaining the context of the entire network. It is important to note the continuity of the lines between the focus and context regions, in spite of the differing magnification factors.



William Farrand's (Farrand 1973) observation that "an effective transformation [of data] must somehow maintain global awareness while providing detail" reflected a longstanding concern, both with a user's need to be aware of context and with the "too much data, too small a screen" problem. Although static solutions already existed in the field of geography, an interactively controlled transformation that satisfied Farrand's requirement and, moreover, maintained a continuity of information space, was invented in 1980 by Robert Spence (Imperial College London) and Mark Apperley (University of Waikato, New Zealand), who gave it the name 'Bifocal Display'. Since then it has been implemented, generalized, evaluated and widely applied. Today there are many applications of the Bifocal Display concept in use; for example the very familiar stretchable dock of application icons associated with the Mac OSX (Modine 2008) operating system (Figure 7.2).


FIGURE 7.2: The very familiar example of the bifocal concept; the Macintosh OSX application 'dock', released in 2001.



VIDEO 7.1: Introduction to the Bifocal Display.

*Courtesy of Rikke Friis Dam and Mads Soegaard. Copyright: CC-Att-ND (Creative Commons Attribution-NoDerivs 3.0 Unported). View full screen or download (544 KB)*


VIDEO 7.2: Main guidelines and future directions.

*Courtesy of Rikke Friis Dam and Mads Soegaard. Copyright: CC-Att-ND (Creative Commons Attribution-NoDerivs 3.0 Unported). View full screen or download (576 KB)*

Video 7.3: How the Bifocal Display was invented and launched.




Video 7.4: The Bifocal Display concept video from 1980.

## 7.1 THE BIFOCAL DISPLAY EXPLAINED

The concept of the Bifocal display can be illustrated by the physical analogy shown in Figures 7.3, 7.4, and 7.5. In Figure 7.3 we see a sheet representing an information space containing many items: documents, sketches, emails and manuscripts are some examples. As presented in Figure 7.3 the information space may be too large to be viewed in its entirety through a window, and scrolling would be needed to examine all information items. However, if the sheet representing the information space is wrapped around two uprights, as in Figure 7.4, and its extremities angled appropriately, a user will see Figure 7.5 part of the information space in its original detail and, in addition, a 'squashed' view of the remainder of the information space. The squashed view may not allow detail to be discerned but, with appropriate encoding (e.g., colour, vertical position) both the presence and the *nature* of items outside the focus region can be interpreted. If an item is noticed in the context region and considered to be potentially of interest, the whole information space can be scrolled by hand to bring that item into detail in the focus region.

Figures 7.3, 7.4, and 7.5 emphasises that the 'stretching' or 'distorting' of information space is central to the concept of the Bifocal Display. The continuity of information space between focus and context regions is a vital feature and especially valuable in the context of map representation (see below).

FIGURE 7.3: An information space containing documents, email, etc.

*Courtesy of Mark D. Apperley and Robert Spence. Copyright: CC-Att-ND-3 (Creative Commons Attribution-NoDerivs 3.0 Unported).*

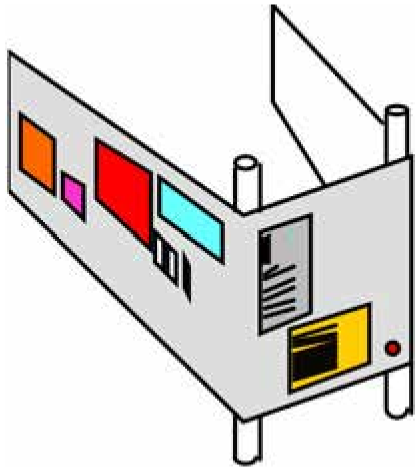


Figure 7.4: The same space wrapped around two uprights.

*Courtesy of Mark D. Apperley and Robert Spence. Copyright: CC-Att-ND-3 (Creative Commons Attribution-NoDerivs 3.0 Unported).*

Figure 7.5: Appearance of the information space when viewed from an appropriate direction.

*Courtesy of Mark D. Apperley and Robert Spence. Copyright: CC-Att-ND-3 (Creative Commons Attribution-NoDerivs 3.0 Unported).*

Immediately following its invention in 1980, the Bifocal Display concept was illustrated in a press release based on an (the first!) envisionment video (Apperley and Spence 1980) showing it in use in the scenario of a futuristic office. It was presented to experts in office automation in 1981 (Apperley and Spence 1981a; Apperley and Spence 1981b;) and the technical details (Apperley et al. 1982) of a potential implementation were discussed in 1982, the same year that a formal journal paper (Spence and Apperley 1982) describing the Bifocal display was published.

A number of significant features of the Bifocal display can be identified:

### 7.1.1 Continuity

Continuity between the focus and context regions in a bifocal representation is an important and powerful feature, facilitated by the notion of 'stretching' or 'distorting' the information space. Formally, the transformation of the space must be monotonic (effectively, moving in the same direction) in both dimensions for continuity to be visible. In fact, the concept of stretching can be generalised. If the stretching shown in Figures 7.5, 7.6, and 7.7 can be termed X-distortion, then stretching in both directions (XY-distortion) can be advantageous in, for example, the display of calendars (Figure 7.6) and metro maps (Figure 7.1): in both these applications the continuity of information space is a distinct advantage. The term 'rubber-sheet stretching' (Tobler 1973; Mackinlay et al. 1991; Sarkar et al. 1993) was seen to neatly explain both the graphical/topological distortion and continuity aspects of focus-plus-context presentations. It is possible that the latter freedom led to use of the term 'fish-eye display' as synonymous with 'bifocal display'. Note that the taxonomy developed by Ying Leung and Apperley (Leung and Apperley 1993a; Leung and Apperley 1993b) discusses the relationships and differences between the bifocal and fish-eye concepts.

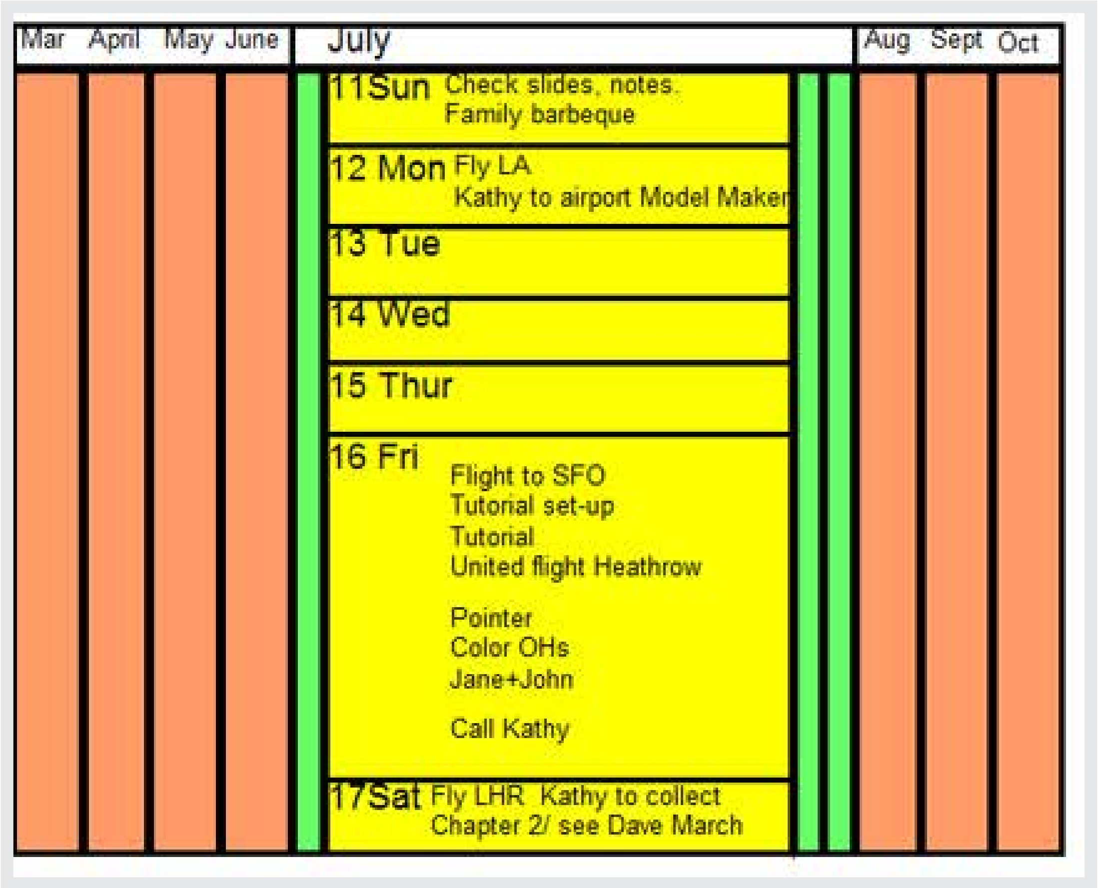



Figure 7.6: Combined X- and Y- distortion provides a convenient calendar interface.



### 7.1.2 Detail Suppression

A second significant feature of the bifocal display is the ability to customise the representation of an item for its appearance in the context region, where fine detail is irrelevant or even inappropriate (see, for example, the London Underground map of Figure 7.1, where no attempt is made to provide station detail in the context region). The concept of 'degree of interest', later to be formalised by George Furnas (Furnas 1986) might, for example lead to the suppression of text and the possible introduction of alternative visual cues, such as shape and colour, with a view to rendering the item more easily distinguished when in the context region. Whereas the bifocal concept is primarily explained as a presentation technique, it was immediately apparent that the effectiveness of the presentations could be enhanced by corresponding variations in representation, utilising the implicit degree of interest of the focus and context regions.

### 7.1.3 Interaction: scrolling/panning

Yet a third feature of the bifocal concept concerned manual interaction with the display to achieve scrolling or panning. In the envisionment video (Apperley and Spence 1980) the user is seen scrolling by touch, immediate visual feedback ensuring easy positioning of a desired item in the focus region (see Figure 7.7). Truly direct manipulation, as in touch, is vital for predictable navigation in a distorted space, and overcomes the issues of scale and speed (Guiard and Beaudouin-Lafon 2004) typically associated with combined panning and zooming operations. The impact and potential of multi-touch interfaces in such interaction is mentioned later.

Figure 7.7: Direct interaction with the Bifocal Display allows a specific item or area to be dragged into the focus region (from Video 5).

*Courtesy of Robert Spence, with the assistance of Colin Grimshaw of the Imperial College TV studio.*

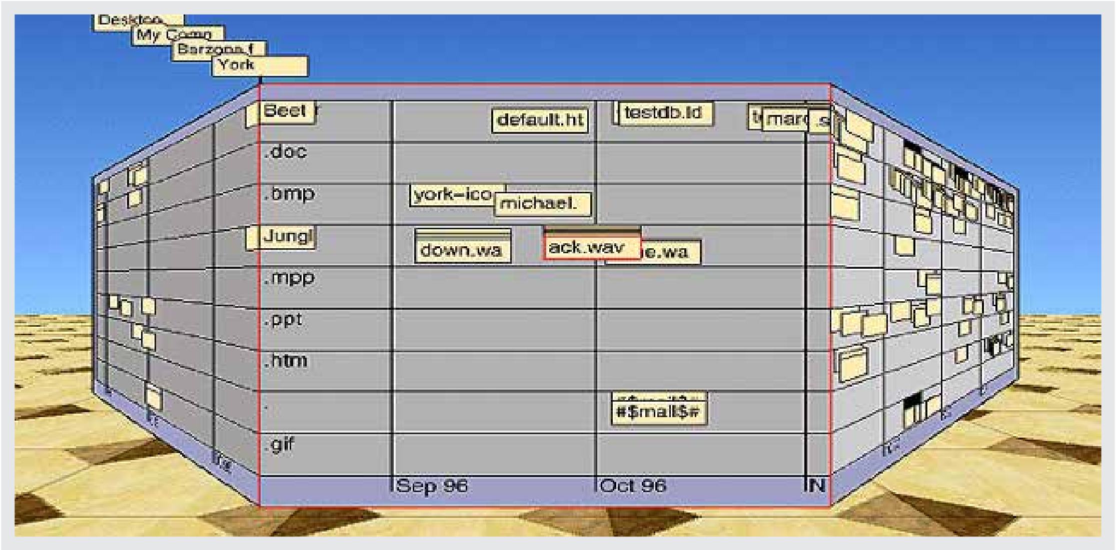



Figure 7.8: The Perspective Wall from 1991 has much in common with the bifocal display.

*Courtesy of Inxight Software, Inc (screenshot of Perspective Wall). Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.*

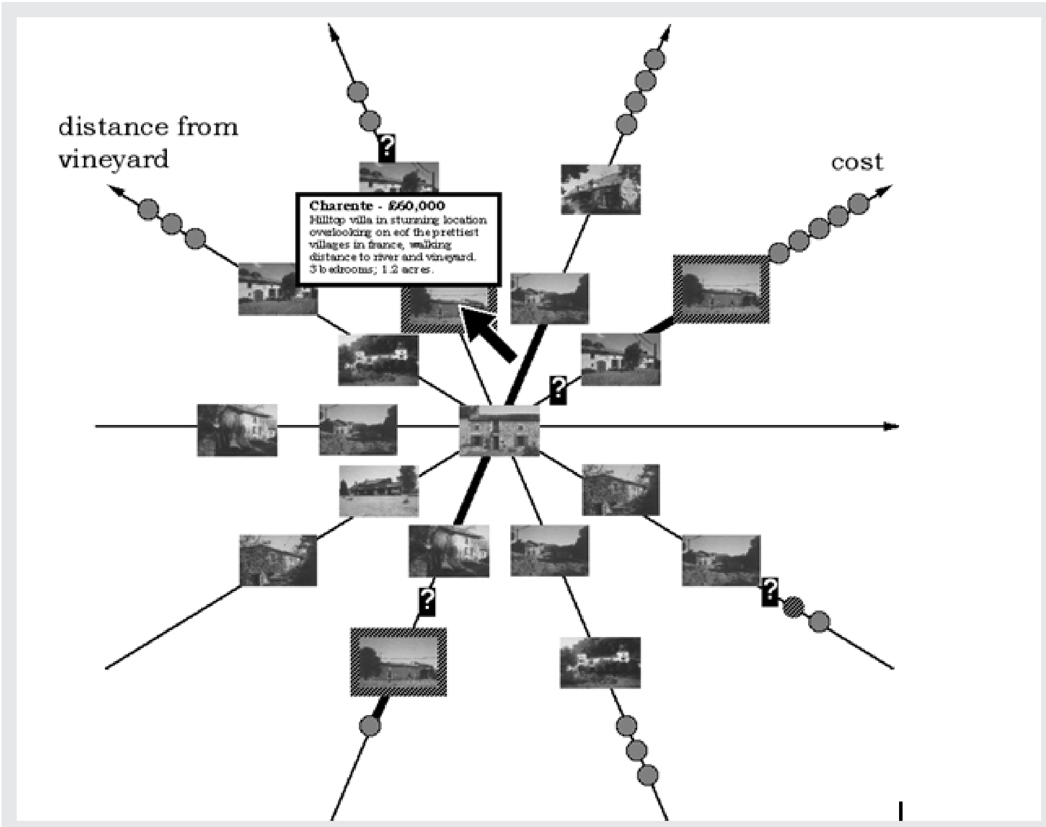



FIGURE 7.9: The Neighbourhood Explorer (Spence 2001; Apperley et al. 2001). Properties further away from the object of interest on each axis are shown as icons with little detail.

*Courtesy of Mark D. Apperley and Robert Spence. Copyright: CC-Att-ND (Creative Commons Attribution-NoDerivs 3.0 Unported).*

Later work by Apperley and Spence and colleagues described generalizations of the Bifocal Display concept and a useful taxonomy (Leung and Apperley 1993a,b,c,d; Leung et al. 1995). In 1991 a three-dimensional realization of the Bifocal Display, termed the Perspective Wall (Figure 7.8), was described (Mackinlay et al. 1991). In the Neighbourhood Explorer (Figure 7.9), Apperley and Spence applied the Bifocal Display concept to the task of home-finding (Spence 2001, page 85; Apperley et al. 2001) in a multi-axis representation. A very effective application of the Bifocal concept to interaction with hierarchically structured data was described by John Lamping and Ramana Rao (Lamping and Rao 1994) who employed a hyperbolic transformation to ensure that, theoretically, an entire tree was mapped to a display (Figure 7.10). In the same year, Rao and Stuart Card (Rao and Card 1994) described the Table Lens (Figure 7.12) which, also, employed the concept of stretching.

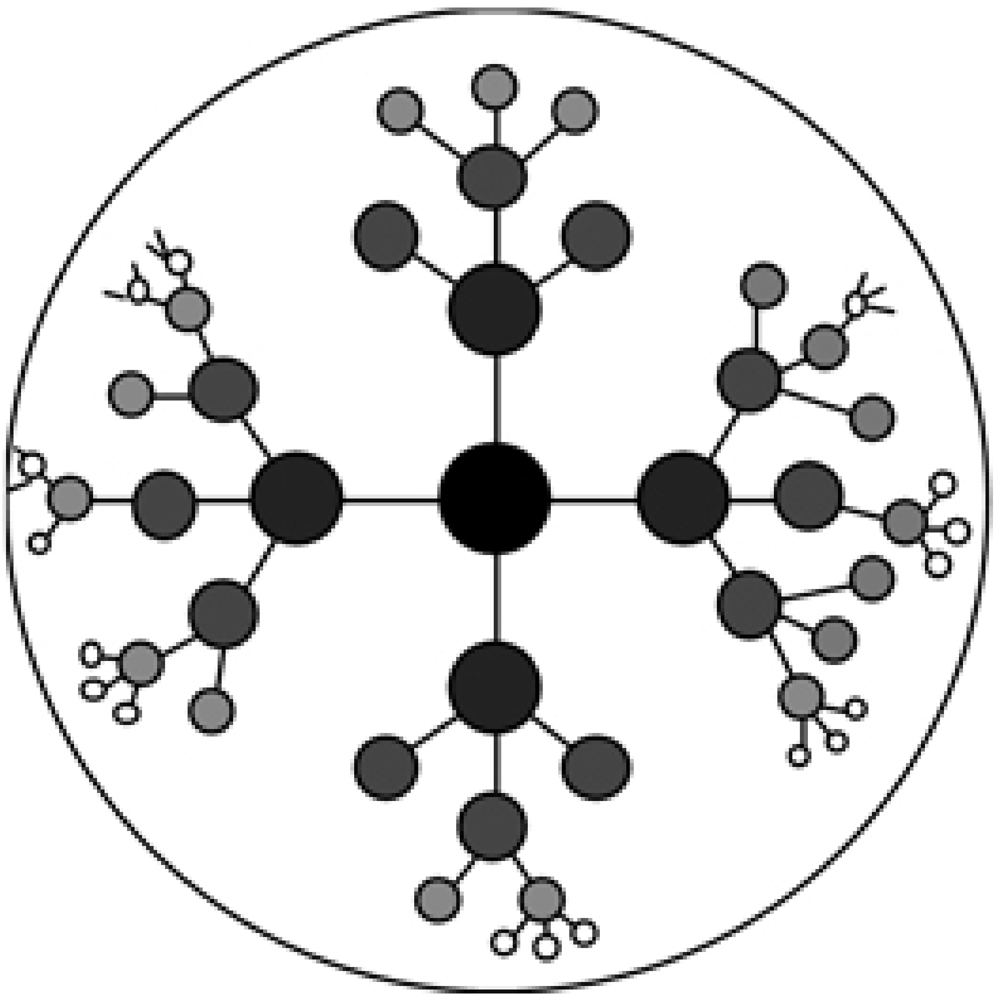


FIGURE 7.10: A sketch illustration of the hyperbolic browser representation of a tree. The further away a node is from the root node, the closer it is to its superordinate node, and the area it occupies decreases (Spence 2001).

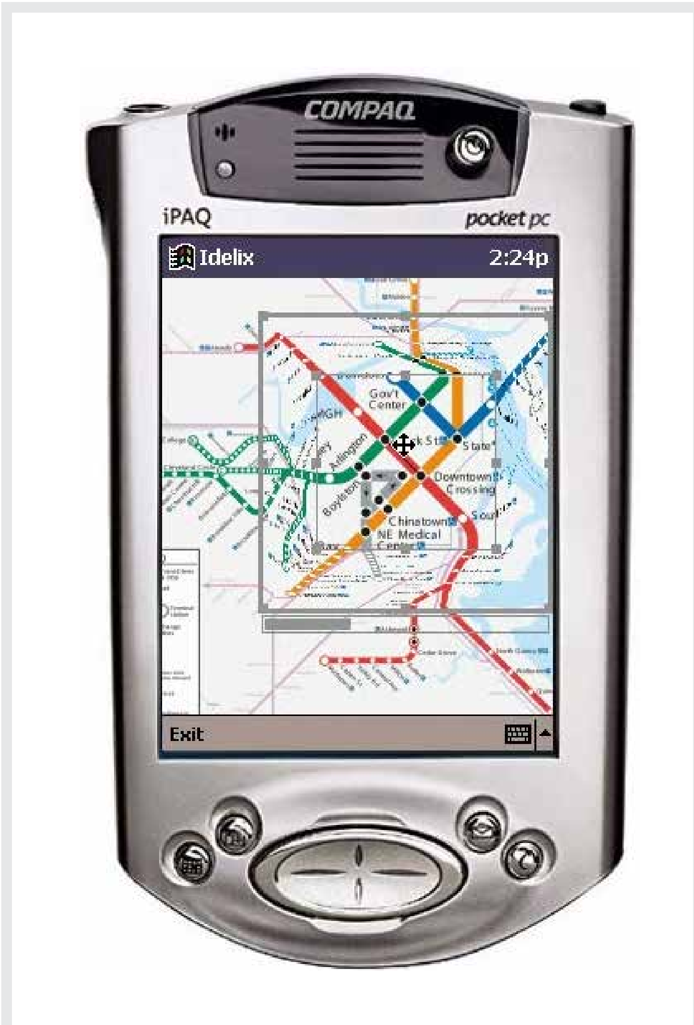



Figure 7.11: Distorted map on a PDA, showing the continuity of transportation links.

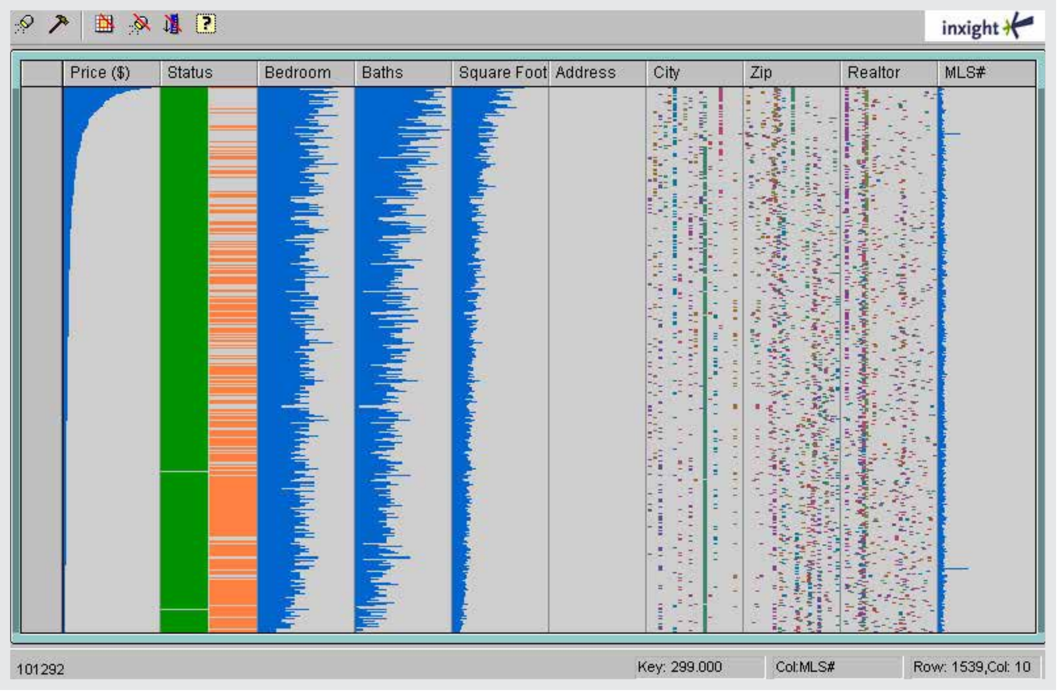



Figure 7.12: Screenshot of the Table Lens. The Table Lens incorporates the concept of stretching in both X and Y dimensions to provide focus plus context (Rao and Card 1994).



The commercial development by IDELIX of software that would implement the concept of the Bifocal Display allowed that company to demonstrate the concept in a number of applications. In one, a transportation map of the Boston area could be examined on the limited display area of a PDA (Figure 7.11) through the appropriate manual control of panning and variable stretching; automatic degree-of-interest adjustment was employed to make the best use of available display area. By contrast, another application (Figures 7.13 and 7.14) employed a table-top display, with four simultaneous users independently controlling the stretching of different areas of the map in order to inspect detail. The value of the Bifocal Display concept to a user's interaction with a calendar was demonstrated by Ben Bederson, Aaron Clamage, Mary Czerwinski and George Robertson (Bederson et al 2004) - see Figure 7.15.

In a medical application of the bifocal concept a 3D image of a portion of the brain has been distorted to focus on the region around an aneurysm, with the surrounding network of arteries as the context (Cohen et al. 2005) - see Figure 7.16 and Figure 7.17.

Figure 7.13: Distorted map on a table (from 2005).

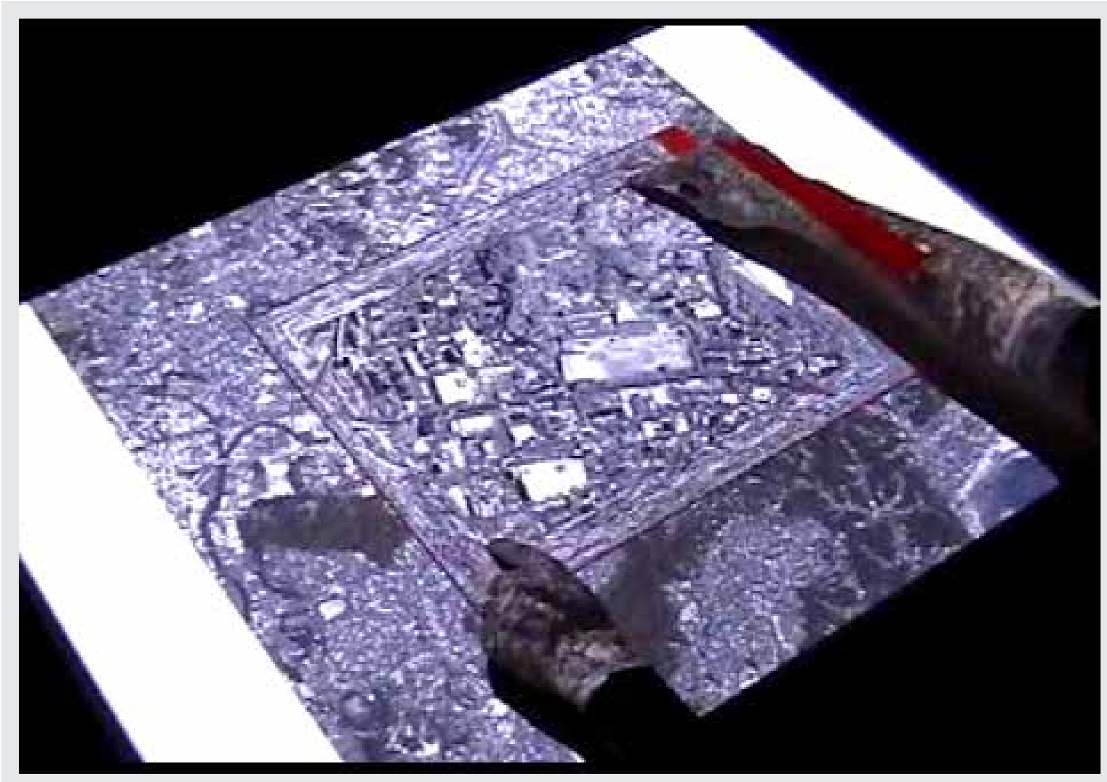


FIGURE 7.14: Distorted map on a table (from 2005).

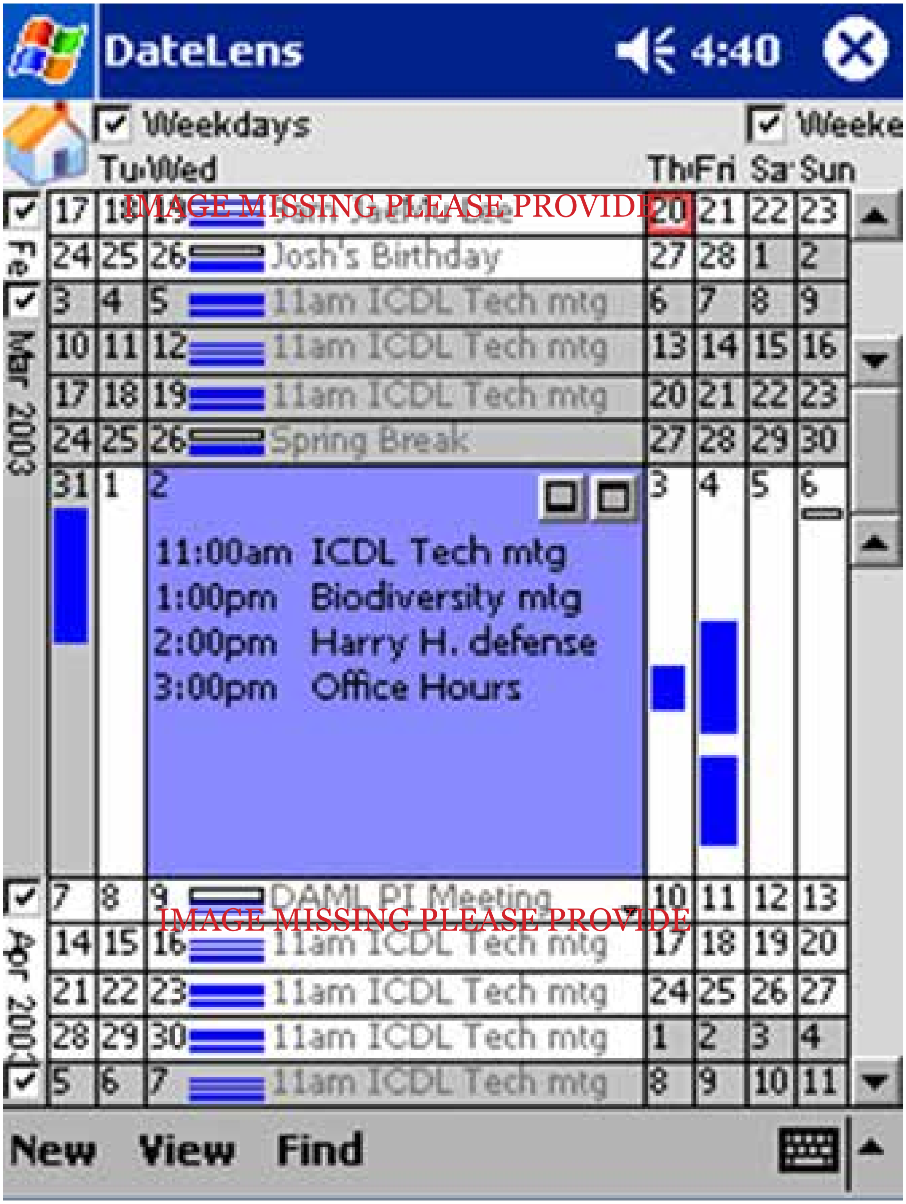



FIGURE 7.15: Use of the Bifocal Display concept in a PDA-based calendar (Bederson et al. 2004).

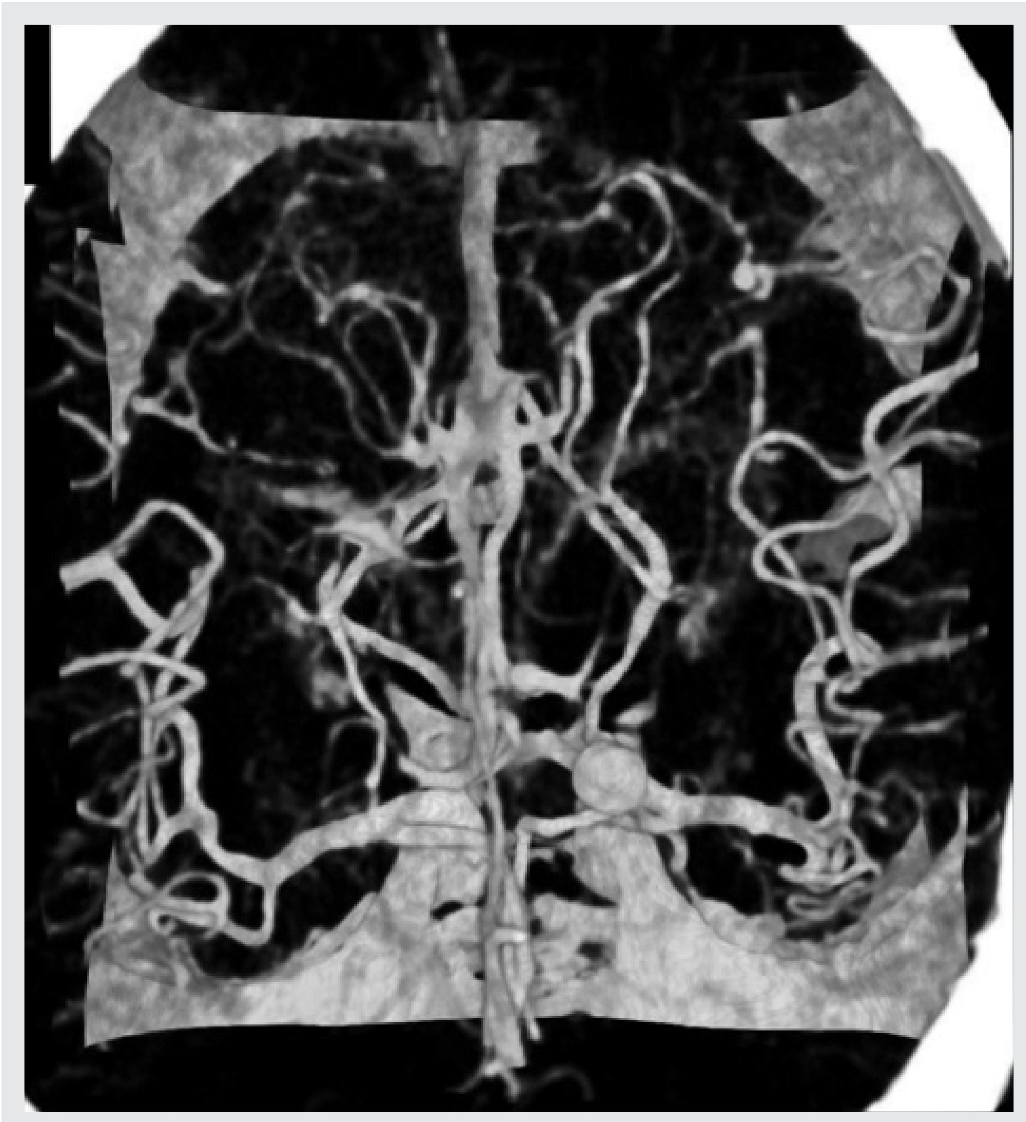


Figure 7.16: A 3D medical dataset of a brain aneurysm without bifocal distortion (Cohen et al. 2005).

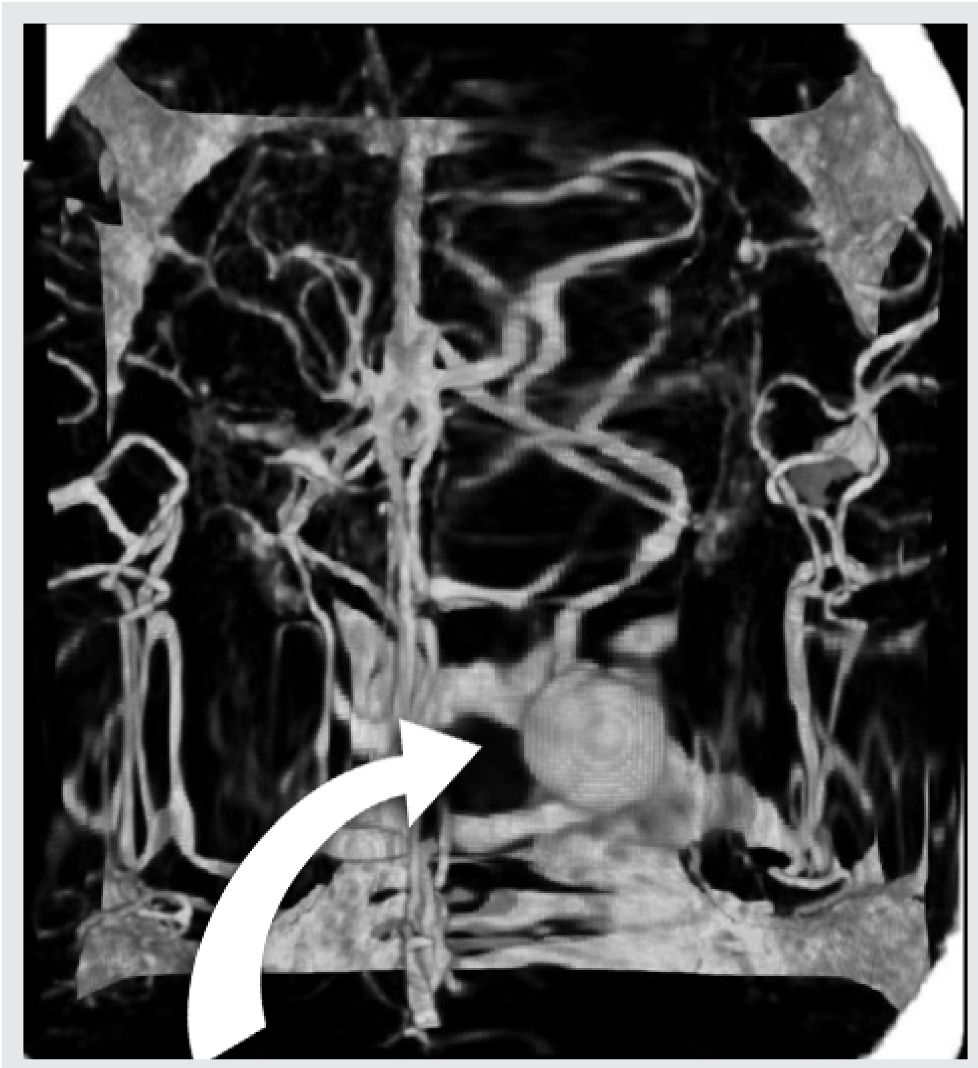


FIGURE 7.17: Bifocal distortion applied to the dataset (Cohen et al. 2005).



## 7.2 THE FUTURE

Research is needed into the fundamental cognitive and perceptual reasons why, and in what circumstances, awareness of context is particularly useful, so that the potential of the bifocal, Degree-of-Interest and other focus+context techniques, alone or in concert, can be assessed for a specific application. The advent of multi-touch screens, and their associated (extreme) direct manipulation, has opened enormous opportunities for improved interaction techniques in navigating large spaces. The single gesture combined pan-zoom operation possible with a multi-touch display offers exciting possibilities for further development and utilisation of the bifocal concept (Forlines and Shen 2005).

## 7.3 WHERE TO LEARN MORE

A chapter of Bill Buxton's book (Buxton 2007) is devoted to the Bifocal Display. The bifocal concept is also treated in many texts associated with Human-computer Interaction, under a variety of index terms: distortion (Ware 2007), bifocal display (Spence 2007; Mazza 2009), and focus+context Tidwell (Tidwell 2005).

## 7.4 VIDEOS

Appreciation of the Bifocal Display concept can be helped by viewing video presentations. A selection is given below.

Video 7.5: The Bifocal Display.

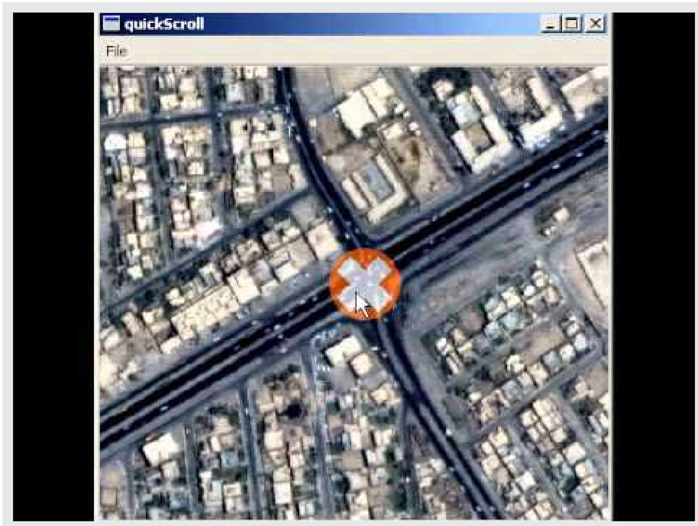


Video 7.6: The Bifocal Display.


Video 7.7: Distorted map on a PDA (52 seconds, silent).

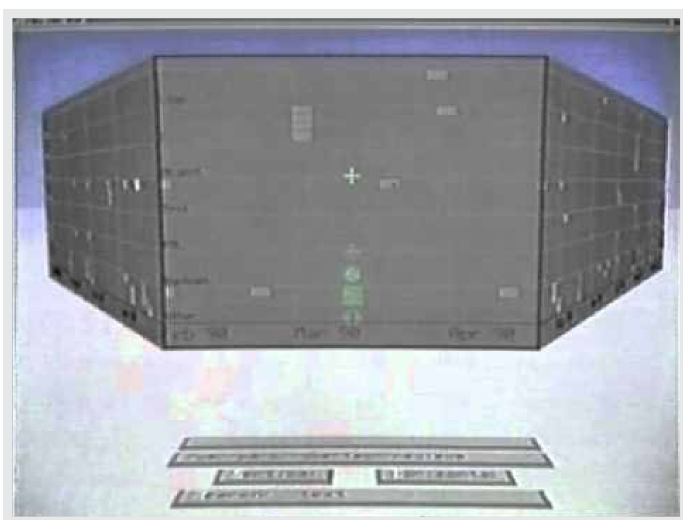


Video 7.8: Pliable display Technology on a table (3 minutes).




Video 7.9: Rubber sheet map distortion (33 seconds, silent).



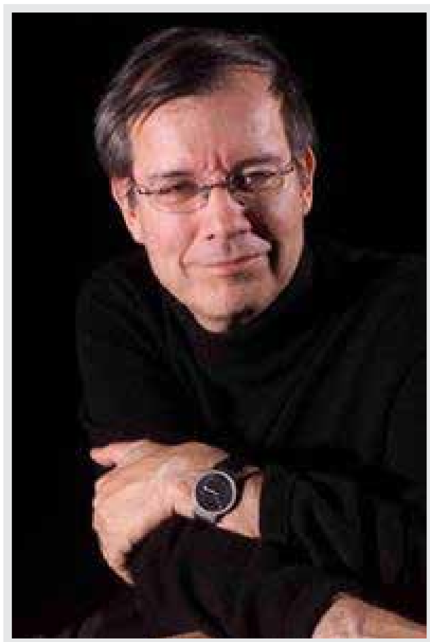


Video 7.10: The Perspective Wall (54 seconds).

## 7.6 COMMENTARY BY STUART K. CARD

How to cite this commentary in your report

### Stuart K. Card



*© Stuart K. Card*

Stuart Card is a Senior Research Fellow and the manager of the User Interface Research group at the Palo Alto Research Center. His study of input devices led to the Fitts's Law characterization of the mouse and was a major factor leading to the mouse's commercial introduction by Xerox. His group has developed theoretical characterizations of human-machine interaction, including the Model...

Stuart K. Card
Stuart K. Card is a member of The Interaction Design Foundation

### 7.6.0.1 The Design Space of Focus + Context Displays

Robert Spence and Mark Apperley have done a fine job of introducing the bifocal display and subsequent explorations of this idea. In this commentary, I want to bring forward the structure of the design space that has emerged and capture some of the abstractions. Then I want to offer a few conjectures about what we have learned about focus + context displays.

The bifocal display is an approach to a general problem: The world presents more information than is possible for a person, with her limited processing bandwidth, to process. A pragmatic solution to this problem is expressed by Resnikoff's (1987) principle of the "selective omission and recoding of information"—some information is ignored while other information is re-encoded into more compact and normalized forms. The bifocal display exemplifies an instance of this principle by dividing information into two parts: a broad, but simplified, *contextual overview part* and a narrow, but detailed, *focal part*. In the contextual overview part, detailed information is ignored or recoded into simplified visual form, whereas in the focal part, more details are included, possibly even enhanced. This roughly mimics the strategy of the human perceptual system, which actually uses a three-level hierarchical organization of retina, fovea, and periphery to partition limited bandwidth between the conflicting needs for both high spacial resolution and wide aperture in sensing the visual environment (Resnikoff, 1987). Visual features picked up in the periphery (for example, a moving something) direct the aim-able, high-resolution fovea/retina and attention to that place of interest, thereby resolving it (for example, into a charging lion).

Spence and Apperley at Imperial College London had the idea that this principle of focus + context could be applied not just to the perceiving agent, but also to the display of the data itself. The working problem for Spence and Apperley was how to organize the dynamic visualization of an electronic workspace. In their solution, documents or journal articles in the focal part were rendered in detail, whereas the documents in the contextual part were foreshortened or otherwise aggregated to take less space and show less detail (Figure 7.1). The detail part of the display could be refocused around a different place in the context area, making it the new focus. Spence and Apperley's method provided a dynamic solution to the use of limited screen space, reminiscent of the dynamics of a pair of bifocal glasses, hence the name *bifocal display*. Their contribution was the conceptual model of the bifocal display, how by using this technique workspaces could be made effectively larger and more efficient, and how this technique could be applied to a broader set of tasks. The first documentation of their technique was expressed in a video of the concept shot in December 1980 (edited in January 1981). Documentation was further published in a journal article in 1982 (Spence and Apperley, 1982)

*Courtesy of Mark D. Apperley and Robert Spence. Copyright: CC-Att-ND-3 (Creative Commons Attribution-NoDerivs 3.0 Unported).*

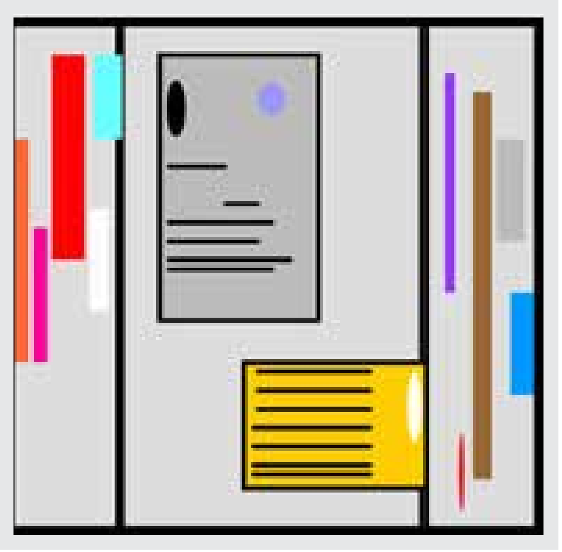




FIGURE 7.1 A-B: Bifocal display applied to desktop workspace (from Figures 7.3 and 7.5 of Spence and Apperley's article). a) Workspace b) Bifocal representation of workspace.

About the same time, George Furnas at Bell Labs had a related idea. Furnas's working problem was how to access statements in long computer program listings. The programmer needed to be able to see lines of code in context, for example declarations of variables that might be several pages back from the current point of interest in the code. He noted that there were intriguing responses to this problem found in everyday life. One famous example is the Steinberg cartoon *New Yorker Magazine* cover showing the world as perceived by a New Yorker on 9th Avenue. Here, detail falls off with increasing distance from 9th Avenue, but there is also more detail than would be expected for Las Vegas and a few other spots of interest to a 9th-Avenue New Yorker. Another example from everyday life is the fisheye lens for a photographic camera, with its distorted enlargement of the central image and shrunken rendering of the image periphery. Furnas's contribution was the invention of a computational degree-of-interest (DOI) function for dynamically assigning a user's relative degree of interest for different parts of a data structure. He then was able to use his DOI function to partition information into more focal and more peripheral parts. His function had two terms, one term expressing the intrinsic importance of something, the other expressing the effect of distance from the point of interest. This function in many cases seemed to create a natural way of compressing information. For example, Figure 7.2, taken from his original 1982 memo, gives a fragment of a computer program when the user's focus is at line 39. After computing the Degree of Interest Function value for each line of the program, those lines with DOI below a threshold are filtered out, resulting in the more compact *fisheye view* in Figure 7.3. The fisheye view version makes better use of space for the program listing. It brings into the listing space information that is at this moment highly relevant to the programmer, such as the includes statement, the variables declaration statement, the controlling while-loop statement, and the conditional statement. It makes room for these by omitting details relevant to the programmer at the moment, such as in some of the case statements. The first documentation of his technique was an internal Bell Labs memo in October of 1982 (Furnas, 1982), widely circulated at the time among the research community, but not formally published until 1999 (Furnas, 1982/1999). The first formal published paper was (Furnas G. , 1986).

```
  28                  t[0] = (t[0] + 10000)
  29                             - x[0];
  30                  for(i=1;i<k;i++){
  31                  t[i] = (t[i] + 10000)
  32                             - x[i]
  33                             - (1 - t[i-1]/10000);
  34                  t[i-1] %= 10000;
  35                  }
  36                  t[k-1] %= 10000;
  37                  break;
  38          case 'e':
>>39                  for(i=0;i<k;i++) t[i] = x[i];
  40                  break;
  41          case 'q':
  42                  exit(0);
  43          default:
  44                  noprint = 1;
  45                  break;
  46  }
  47  if(!noprint){
  48          for(i=k - 1;t[i] <= 0 && i &rt; 0;i--);
  49          printf("%d",t[i]);
  50          if(i &rt; 0) {
```

FIGURE 7.2: Fragment of program listing before applying fisheye view. Fisheye view of program listing. Line 39 (in red) is the focus.

```
1 #define DIG 40
2 #include
...4 main()
5 {
6 int c, i, x[DIG/4], t[DIG/4], k = DIG/4, noprint = 0;
...8 while((c=getchar()) != EOF){
9 if(c >= '0' && c <= '9'){
...16 } else {
17 switch(c){
18 case '+':
...27 case '-':
...38 case 'e':
>>39 for(i=0;i<k;i++) t[i] = x[i];
40 break;
41 case 'q':
...43 default:
...46 }
47 if(!noprint){
...57 }
58 }
59 noprint = 0;
60 }
61 }
```

Figure 7.3: A fisheye view of the C program. Line numbers are in the left margin. "..." indicates missing lines. Note that the variable declarations and while-loop initiation are now on the same page. Line 39 (in red) is the focus.

It is helpful to consider bifocal displays or fisheye views as contrasted with an alternative method of accessing contextual and focal information: *overview + detail*. Figure 7.4 shows the data of the Spence and Apperley bifocal display as an overview + detail display. The advantage of overview + detail is that it is straightforward; the disadvantage is that it requires moving the eye back and forth between two different displays. The bifocal display essentially seeks to fit the detail display within the contextual display, thereby avoiding this coordination and its implied visual search.

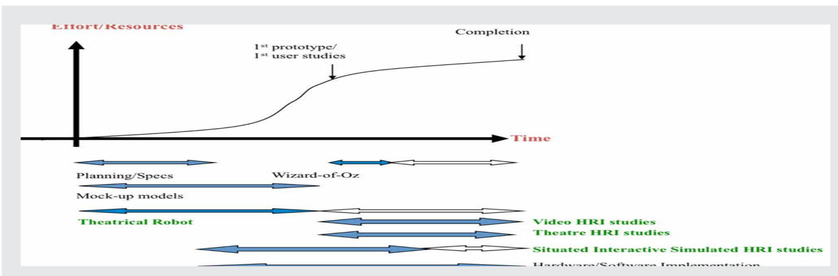







Figure 7.4 A-B: Overview + detail display. a) Overview b) Detail.

Despite the original names of "bifocal display" or "fisheye view", the collection of techniques derived from these seminal papers, both by the authors themselves as well as by others, go well beyond the visual transformation fisheye implies and beyond two levels of representation the name bifocal implies. These displays might be called *attention-aware displays* because of the way in which they use proxies for user attention to dynamically reallocate display space and detail. Pragmatically, I will refer to the general class as *focus + context* techniques to emphasize the connection beyond the visual to user attention and to avoid having to say "bifocal display or fisheye view" repeatedly.

#### 7.6.0.2 Focus + Context Displays as Visualization Transformations

Focus + context techniques are inherently dynamic and lead us to think of information displays in terms of space × time × representation transformations. The classes of representations available can be seen in terms of the *information visualization reference model* (Card, Mackinlay, & Shneiderman, 1999) reproduced in Figure 7.5. This framework traces the path from raw data to visualization as the data is transformed to a normalized form, then mapped into visual structures, and then remapped into derivative visual forms. The lower arrows in the diagram depict the fact that information visualizations are dynamic. The user may alter the parameters of the transformations for the visualizations she is presently viewing

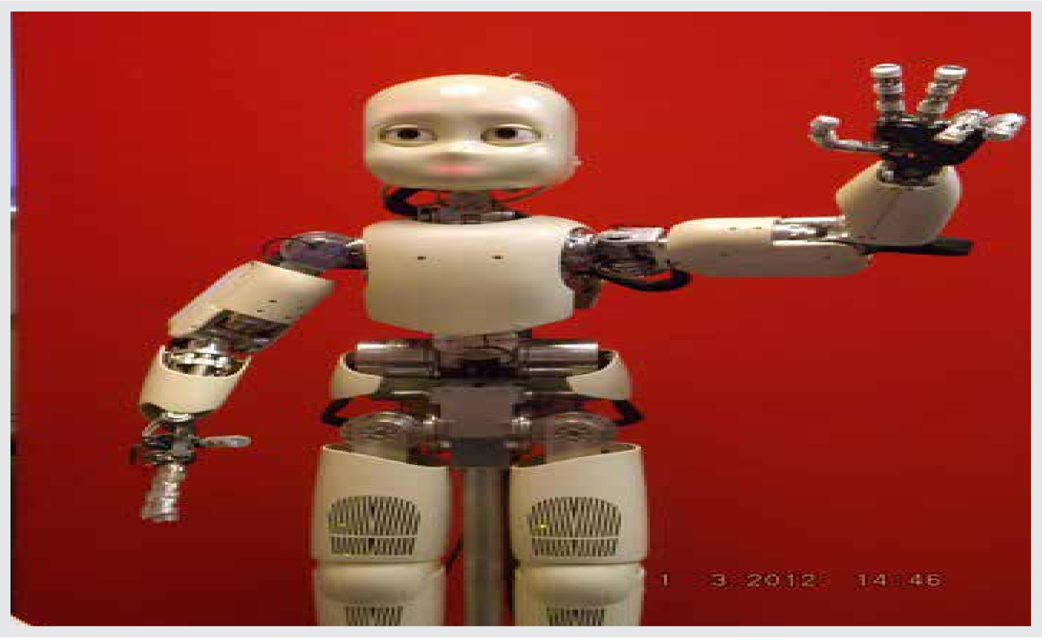



FIGURE 7.5: Information visualization reference model (Card, Mackinlay, and Shneiderman, 1999).



Focus + context displays mix the effects of two transformations of the Information Visualization Reference Model: *view transformations* and *visual mappings*. View transformations use a mapping from space into space that distorts the visualization in some way. Some can be conveniently described in terms of a *visual transfer function* for achieving the focus + context effect. The bifocal display was the first of these and inspired later work.

Visual mappings are concerned with a mapping from data to visual representation, including filtering out lower levels of detail. The design space of filters for visual mappings with respect to filtering can often be conveniently described in terms of choices for *degree-of-interest functions* applied to the structure or content of the data and how these are used to filter level of detail. It also inspired later work.

This convenient historical correlation, however, between geometrically-oriented techniques and the bifocal display on the one hand and data-oriented level-of-detail filtering degree-of-interest techniques on the other does *not* reach to the essence of these techniques either analytically or historically. Even in the initial papers, Spence and Apperley did not simply apply geometrical transformations, but also understood the advantages of changing the representation of the data in context and focal parts of the display, as in Figure 6 from their original paper, which shows a simple representation of months in the context part of the display expanded to a detailed representation of daily appointments in the focal part of the display. Conversely, Furnas in his first memo on the fisheye view included a section on "Fisheye view of Euclidean space" and so understood the potential use of his technique to visual transformations. Nor do these techniques exhaust the possibilities for dynamic focus + context mappings.

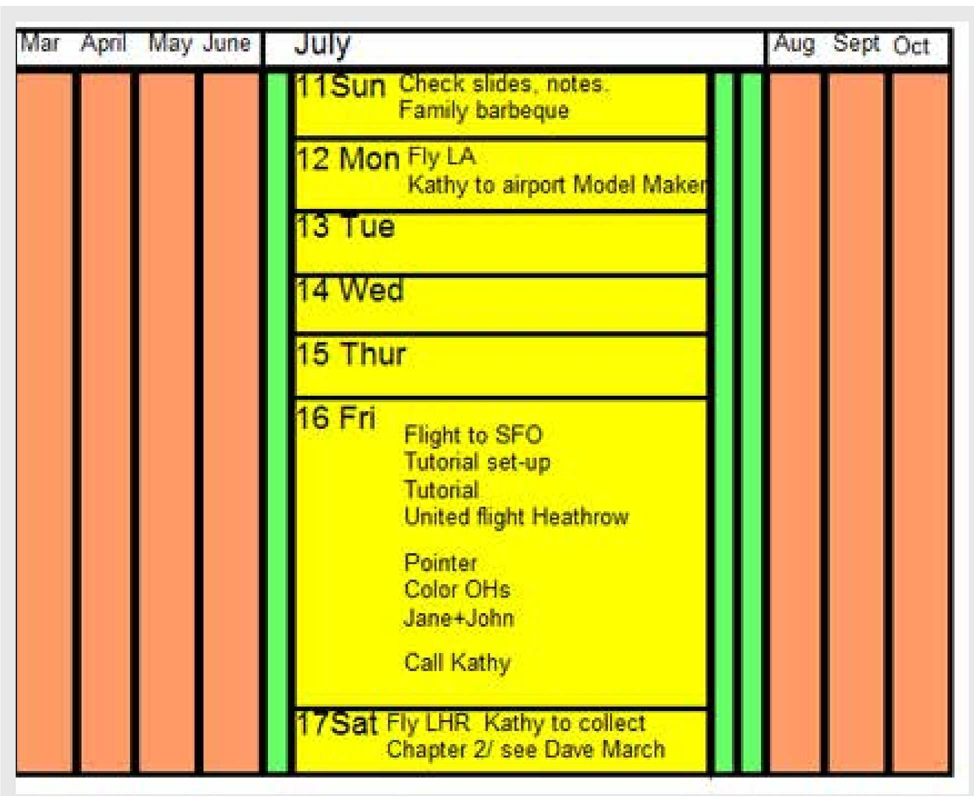



Figure 7.6: Example of bifocal display semantic representation change (from Spence and Apperley's Figure 7.6).

The essence of both bifocal displays and fisheye views is that view transformations and visual mapping transformations actively and continually change the locus of detail on the display to support the task at hand. The combination of possible transformations generates a design space for focus + context displays. To appreciate the richness of this design space generated by the seminal ideas of Spence & Apperley and Furnas, we will look at a few parametric variations of visual transfer functions and degree-of-interest functions.

### 7.6.0.3 View Transformations as Visual Transfer Functions

View transformations transform the geometry of the space. The bifocal display workspace has two levels of magnification, as illustrated in Figure 7.7.B. From the function representing these two levels of magnification, we can derive the visual transfer function in Figure 7.7.C., which shows how a point in the image is transformed. The two levels of constant magnification in the magnification function, one for the peripheral context region, the other for the focal region, yield a visual transfer function (which is essentially the integral of the magnification function). The result of applying this transformation to the original image, Figure 7.7.A., is the image shown in Figure 7.7.D, foreshortening it on the sides.

*Courtesy of Mark D. Apperley and Robert Spence. Copyright: CC-Att-ND-3 (Creative Commons Attribution-NoDerivs 3.0 Unported).*

| | | | | | | | fidelity |
|---|---|---|---|---|---|---|---|
| | +++ | +++ | + | +++ | +++ | - | - |
| I | + | - | ++ | - | ++ | +++ | - |
| I | ++ | - | ++ | - | ++ | +++ | + |
| RI | +++ | - | +++ | +++ | +++ | - | + |
| | +++ | +++ | +++ | +++ | - | - | +++ |

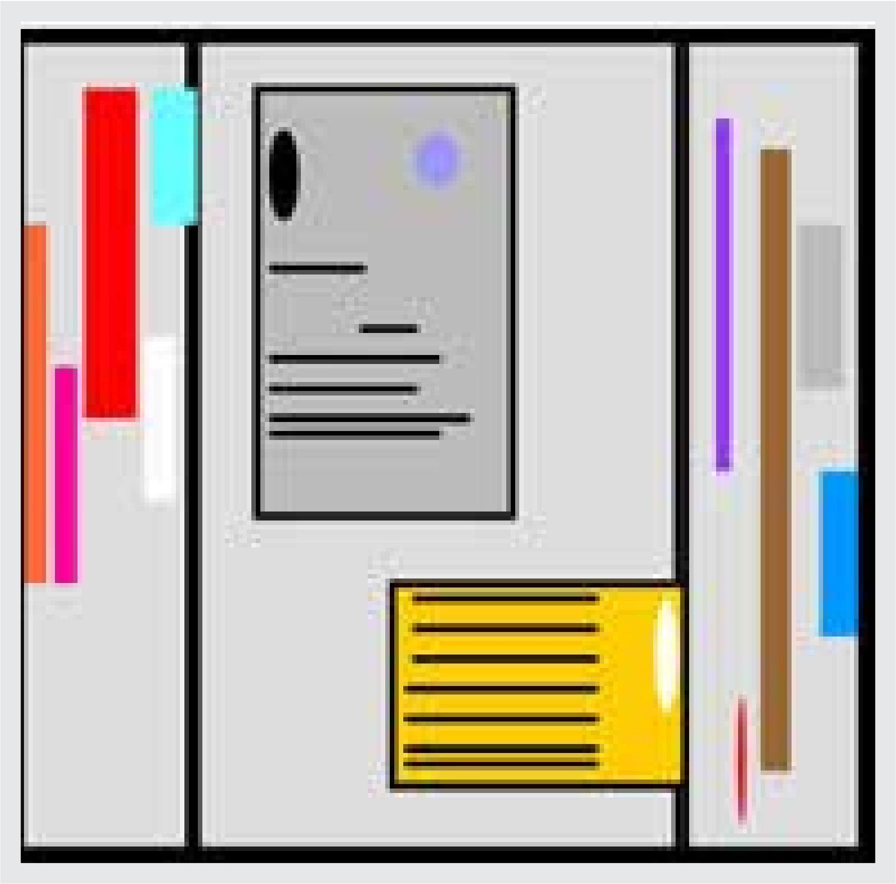




Figure 7.7 A-B-C-D: Bifocal display visual transfer function of a bifocal display.: (a) Original image (b) Magnification function (c) Visual Transfer function (d) Transformed workspace.

*Rubber Geometry: Alternate Visual Transfer Functions*. It is apparent that the visual transfer function can be generalized to give many alternate focus + context displays. Leung and Apperley (1994) realized early on that the visual transfer function was a useful way to catalogue many of the variations of these kinds of displays and did so. Ironically, among the first of these addressed by Leung and Apperley (1994) is the visual transfer function of a true (optical) fisheye lens, which had mostly been discussed metaphorically by Furnas (1982). The fisheye magnification function (Figure 7.8.A) and the resulting visual transfer function (Figure 7.8.B) result in the transformed workspace in Figure 7.8.C, depicted by showing how it distorts gridlines



Figure 7.8 A-B-C: Visual Transfer function of a fisheye lens (Leung and Apperley, 1994): (a) Magnification function. (b) Visual Transfer function. (c) Transformed workspace.

Notice that in Figure 7.8.A rather than just two magnification levels, there is now a con-tinuous function of them. Notice also that unlike Figure 7.7.C, which describes a one-dimensional function, Figure 7.8.B is shorthand for a two-dimensional function, as is apparent in Figure 7.8.B. There are many forms the visual transfer function could take. An interesting subset of them is called rubber sheet transfer functions, so-called because they just seem to stretch a continuous sheet. Figure 7.9 shows a few of these.



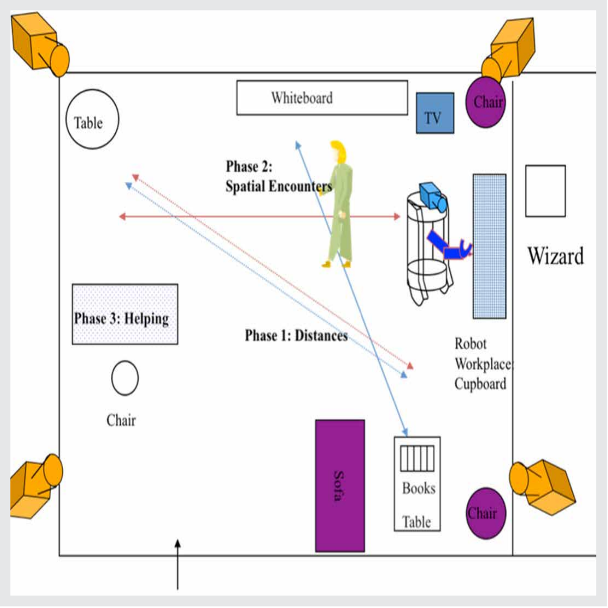


Table
Whiteboard
TV
Chair
Phase 2:
Spatial Encounters
Wizard
Phase 3: Helping
Phase 1: Distances
Robot
Workplace
Cupboard
Chair
Sofa
Books
Table
Chair

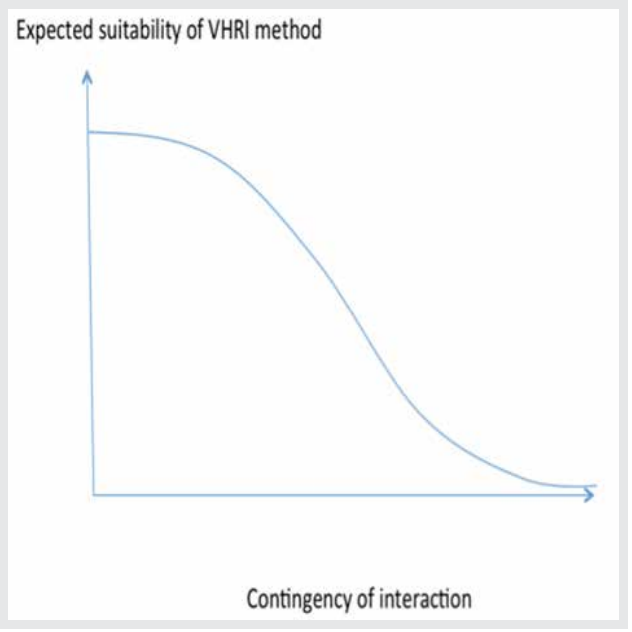





Figure 7.9 A-B-C: Rubber sheet visual transform functions. (Carpendale, 2001). (a) Gausian Transfer function (b) Cosine Transfer function (c) Linear Transfer function.

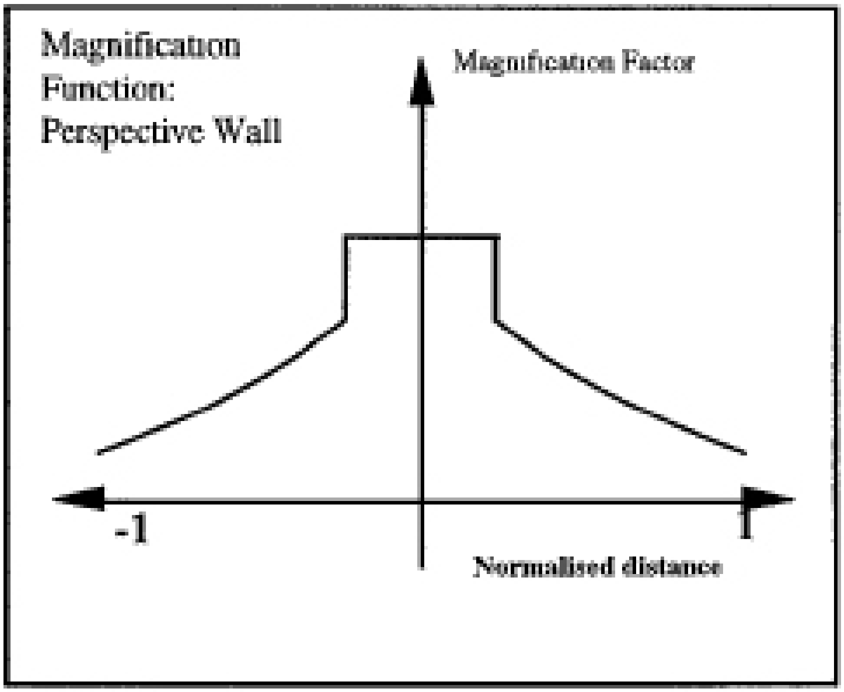
*Natural Perspective Visual Transfer Functions*. One problem with rubber sheet visual transfer functions is that the distortion can be somewhat difficult to interpret, as the mapping from original (Figure 7.10.A) to transformed image (Figure 7.10.B) shows, although this can be mitigated by giving the visual transfer function a flat spot in the center.

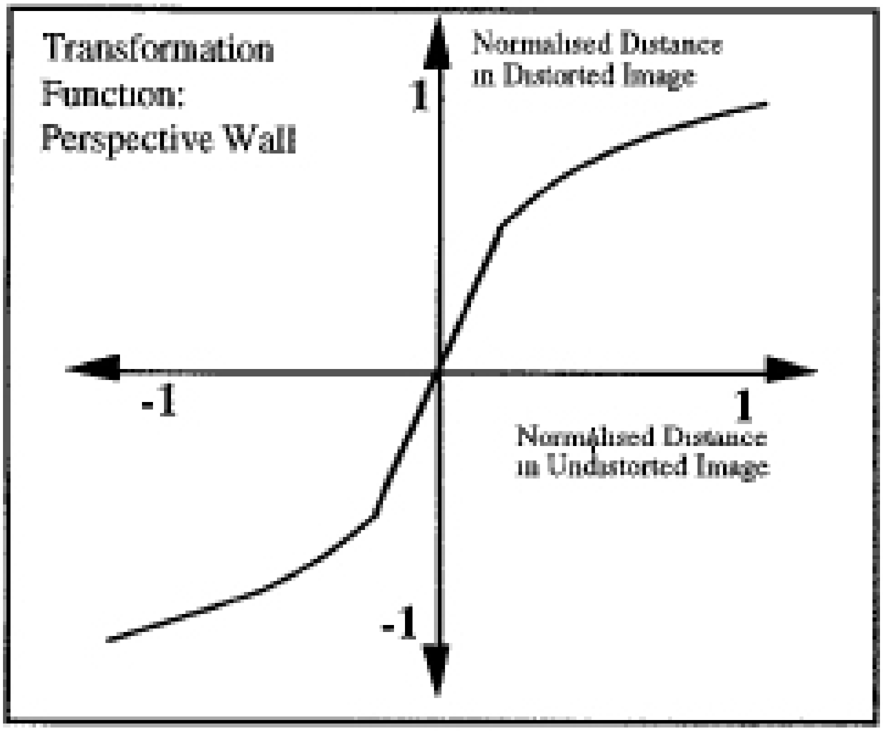




FIGURE 7.10 A-B: Example of distortion engendered by some visual transfer functions (Carpendale, 2006/2012). (a) Original image (b) Transformed image.

An interesting alternative is to use natural perspective visual transfer functions. These functions achieve the required contrast in magnification between the two regions, but the trick is that the display doesn't look distorted. The perspective wall (Figure 7.11.C) is such a display. As we can see by the magnification function (Figure 7.11.A), part of the magnification function is flat, thereby solving the distortion problem, but part of the magnification function on the sides is curved. Yet the curved sides do not appear distorted because the curve matches natural perspective and so is effectively reversed by the viewer's perceptual system (although comparative judgments can still be adversely affected). Touching an element on one of the side panels causes the touched part of the "tape" to slide to the front thereby achieving the magnification of the magnification function in Figure 7.11.A and moving contextual information into focal position. The point is that by using a natural perspective visual transfer function, we get the space-saving aspects of focus + context displays, but the user doesn't think of it as distortion. It just seems natural.


Magnification
Function:
Perspective Wall
Magnification Factor
-1
1
Normalised distance


Transformation
Function:
Perspective Wall
Normalised Distance
in Distorted Image
1
-1
1
Normalised Distance
in Undistorted Image
-1

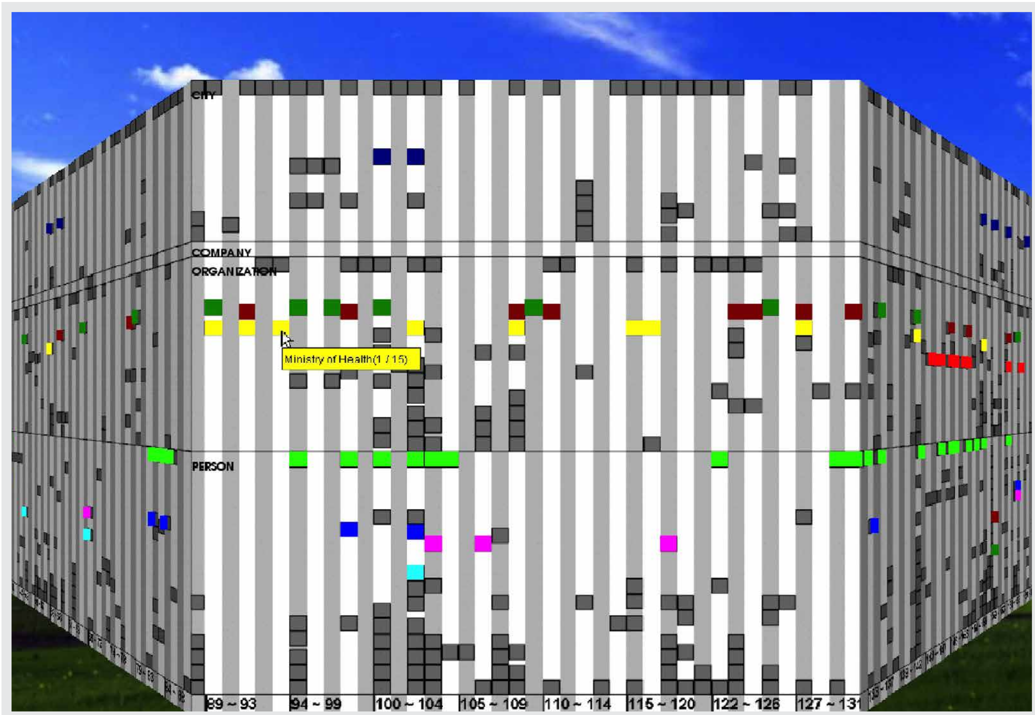





FIGURE 7.11 A-B-C: The perspective wall (Mackinlay, Robertson, and Card, 1991): (a) Magnification function (b) Visual Transfer function (c) Transformed workspace.

*Three-Dimensional Visual Transfer Functions*. The perspective wall introduces another element of variation. The visual transfer function can be in three dimensions. Figure 7.12 shows another such visualization, the document lens (Robertson & Mackinlay, 1993). The document lens is used with a book or a report (Card, Robertson, & York, 1996). The user commands the book to change into a grid of all the book's pages. A search lights up all the phrases of interest and makes clear which pages would be most interesting to examine in detail. The user then reaches in and pulls some pages forward, resulting in Figure 7.12. Even though she is reading one (or a set) of pages in her detail area, all of the pages remain visible as context. Furthermore, since this is a perceptual transformation, the context pages are not experienced as distorted.

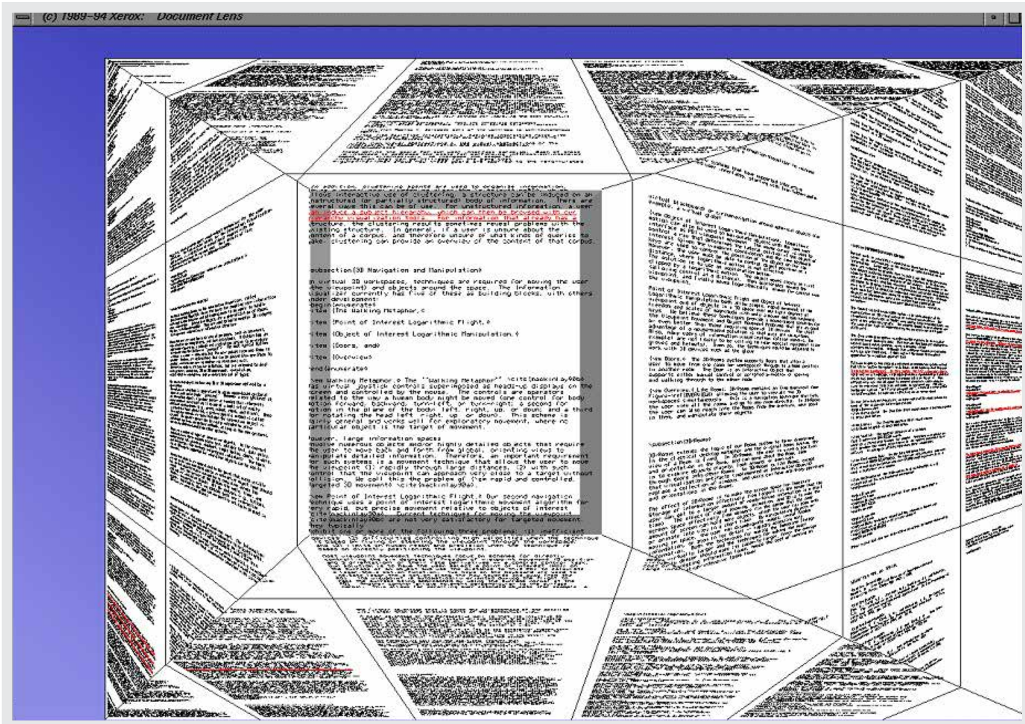
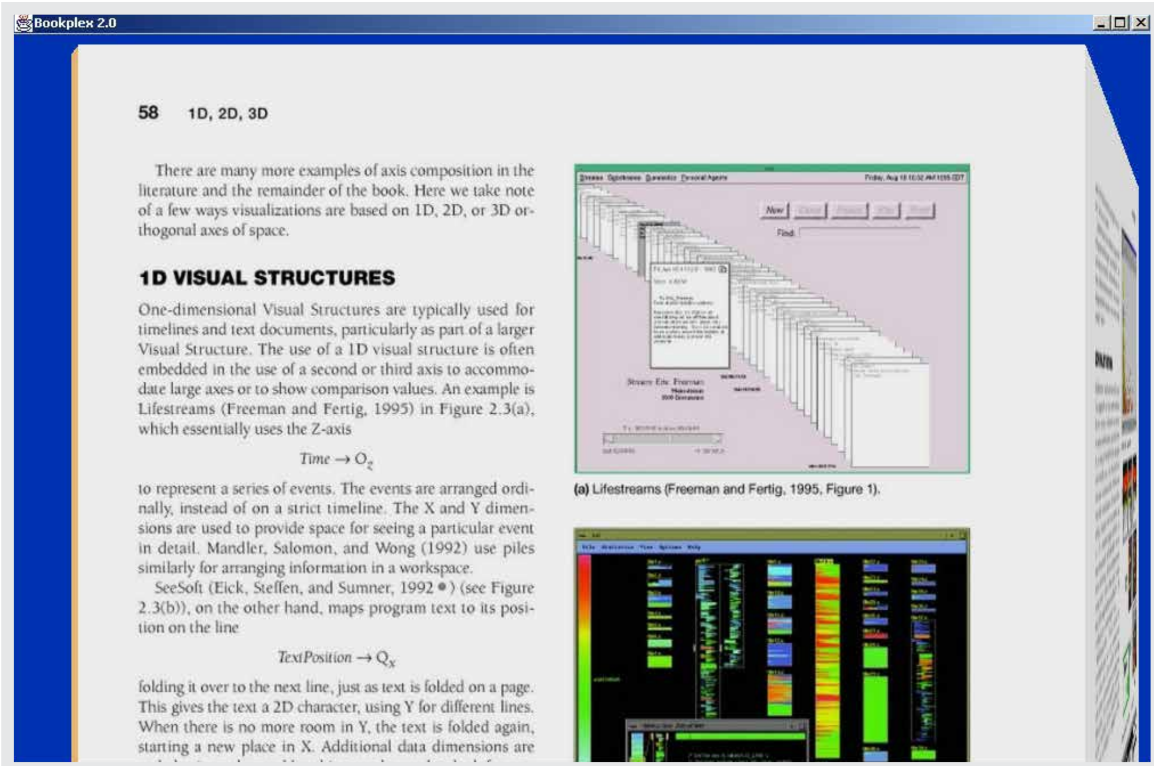



FIGURE 7.12: The Document Lens (Robertson and Mackinlay, 1993).



Natural perspective visual transfer functions fit almost invisibly into strong visual met-aphors and so can be used to produce focus + context effects without drawing attention to themselves as a separate visualization. Figure 7.13.A shows 3Book (Card, Hong and Mackinlay, 2004) a 3D electronic book. There is not room on the screen to show the double page open book, so the view is zoomed into the top left-hand page (the focus) and the right-hand page is bent backward but not completely, so the contents on it are still visible (the context). The reader can see that there is an illustration on the right-hand page and clicking on it causes the book to rock to the position shown in Figure 7.13.B, thus making the right-hand page the focus and the left-hand page the context. In this way, the rocker page focus + context technique is able to preserve more context for the reader while fitting within the available space resource.


Bookplex 2.0

58 1D, 2D, 3D

There are many more examples of axis composition in the literature and the remainder of the book. Here we take note of a few ways visualizations are based on 1D, 2D, or 3D orthogonal axes of space.

**1D VISUAL STRUCTURES**

One-dimensional Visual Structures are typically used for timelines and text documents, particularly as part of a larger Visual Structure. The use of a 1D visual structure is often embedded in the use of a second or third axis to accommodate large axes or to show comparison values. An example is Lifestreams (Freeman and Fertig, 1995) in Figure 2.3(a), which essentially uses the Z-axis

$$Time \rightarrow O_z$$

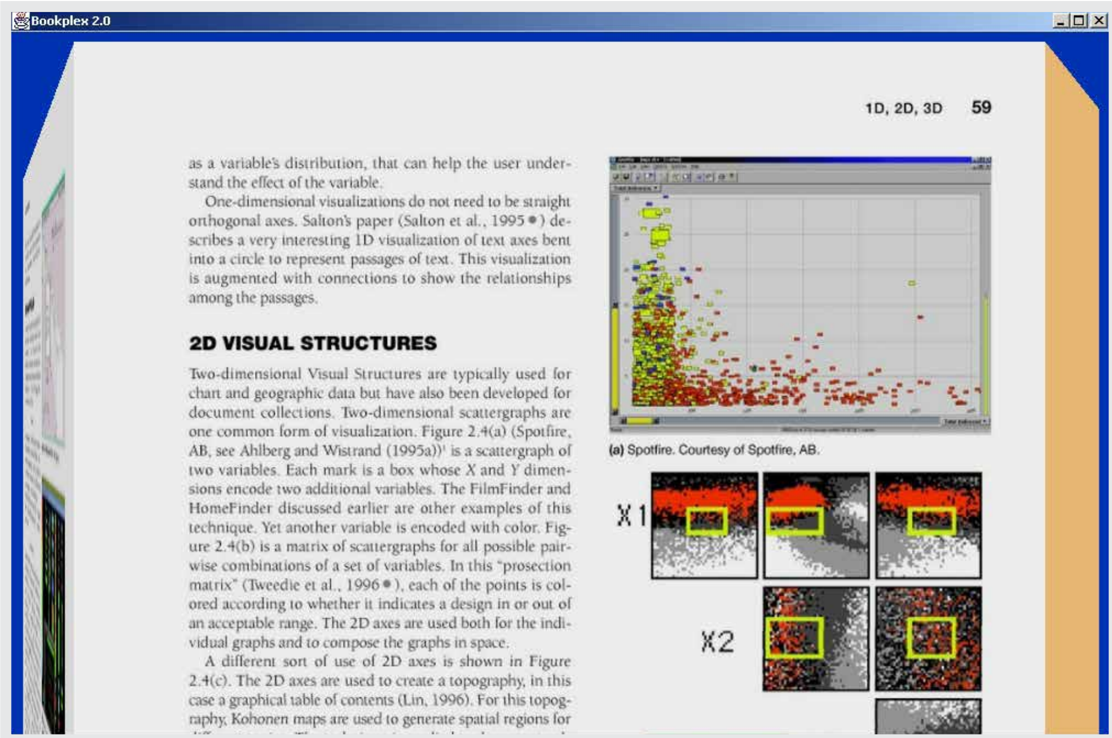
to represent a series of events. The events are arranged ordinally, instead of on a strict timeline. The X and Y dimensions are used to provide space for seeing a particular event in detail. Mandler, Salomon, and Wong (1992) use piles similarly for arranging information in a workspace.

SeeSoft (Eick, Steffen, and Sumner, 1992 ●) (see Figure 2.3(b)), on the other hand, maps program text to its position on the line

$$TextPosition \rightarrow Q_x$$

folding it over to the next line, just as text is folded on a page. This gives the text a 2D character, using Y for different lines. When there is no more room in Y, the text is folded again, starting a new place in X. Additional data dimensions are

(a) Lifestreams (Freeman and Fertig, 1995, Figure 1).



1D, 2D, 3D 59

as a variable's distribution, that can help the user understand the effect of the variable.

One-dimensional visualizations do not need to be straight orthogonal axes. Salton's paper (Salton et al., 1995 ●) describes a very interesting 1D visualization of text axes bent into a circle to represent passages of text. This visualization is augmented with connections to show the relationships among the passages.

**2D VISUAL STRUCTURES**

Two-dimensional Visual Structures are typically used for chart and geographic data but have also been developed for document collections. Two-dimensional scattergraphs are one common form of visualization. Figure 2.4(a) (Spotfire, AB, see Ahlberg and Wistrand (1995a))[1] is a scattergraph of two variables. Each mark is a box whose *X* and *Y* dimensions encode two additional variables. The FilmFinder and HomeFinder discussed earlier are other examples of this technique. Yet another variable is encoded with color. Figure 2.4(b) is a matrix of scattergraphs for all possible pairwise combinations of a set of variables. In this "prosection matrix" (Tweedie et al., 1996 ●), each of the points is colored according to whether it indicates a design in or out of an acceptable range. The 2D axes are used both for the individual graphs and to compose the graphs in space.

A different sort of use of 2D axes is shown in Figure 2.4(c). The 2D axes are used to create a topography, in this case a graphical table of contents (Lin, 1996). For this topography, Kohonen maps are used to generate spatial regions for

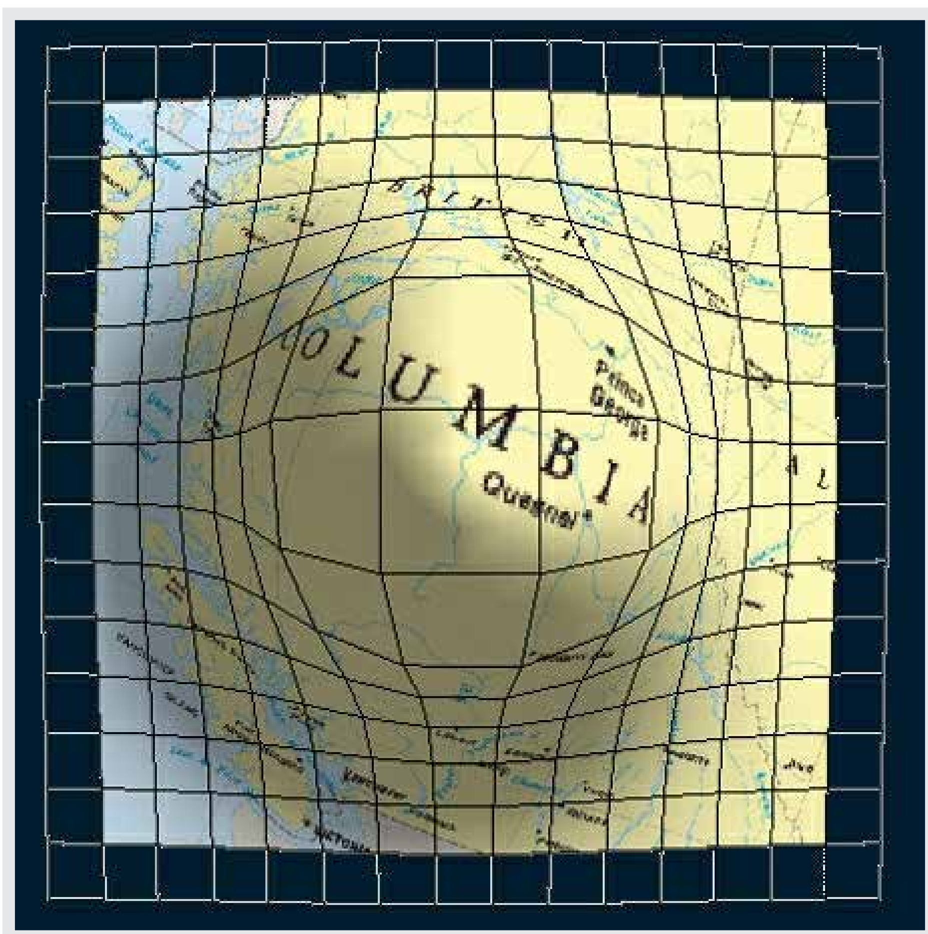


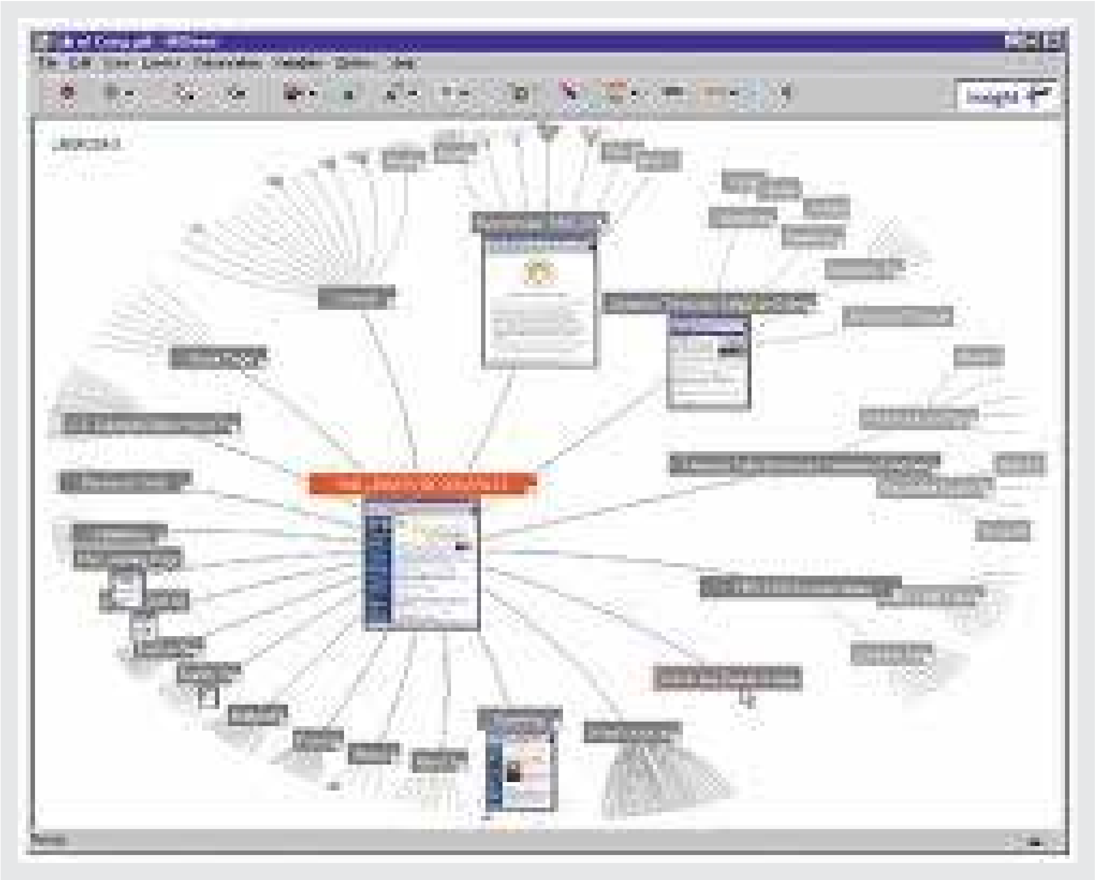


Figure 7.13 A-B: Book use of rocker page focus + context effect (Card, Hong, and Mackinlay, 2004): (a) Left-hand page is focus. Right-hand page is bent partially back, forming context. (b) Book rocks causing left-hand page to become context and right-hand page to become focus.

*Hyperbolic Visual Transfer Functions.* One particularly interesting visual transfer function that has been tried is a hyperbolic mapping. With a hyperbolic function it is possible to compensate for the exponential growth of a graph by shrinking size space on which the graph is projected. This is because an infinite hyperbolic space can be projected onto a finite part of a Euclidean space. As with all focus + context techniques, the part of the graph that is the focus can be moved around with the size adjusted appropriately. Figure 7.14 shows examples of hyperbolic visual transfer functions. Figure 7.14.A is the hyperbolic equivalent to Figure 7.9.

Figure 7.14.B shows a hyperbolic tree (Lamping, Rao, and Pirolli (1995). Notice how the nodes are re-represented as small documents when the space gets large enough. Figure 7.14.C gives a 3D version (Munzner & Burchard, 1995; Munzner, 1998). For fun, Figure 7.14.D shows how this idea could be taken even further using a more extreme hyperbolic projection (in this case, carefully constructed by knitting) (Tallmina, 1997) that could serve as an alternate substrate for trees to that in Figure 7.14.B or 7.14.C).





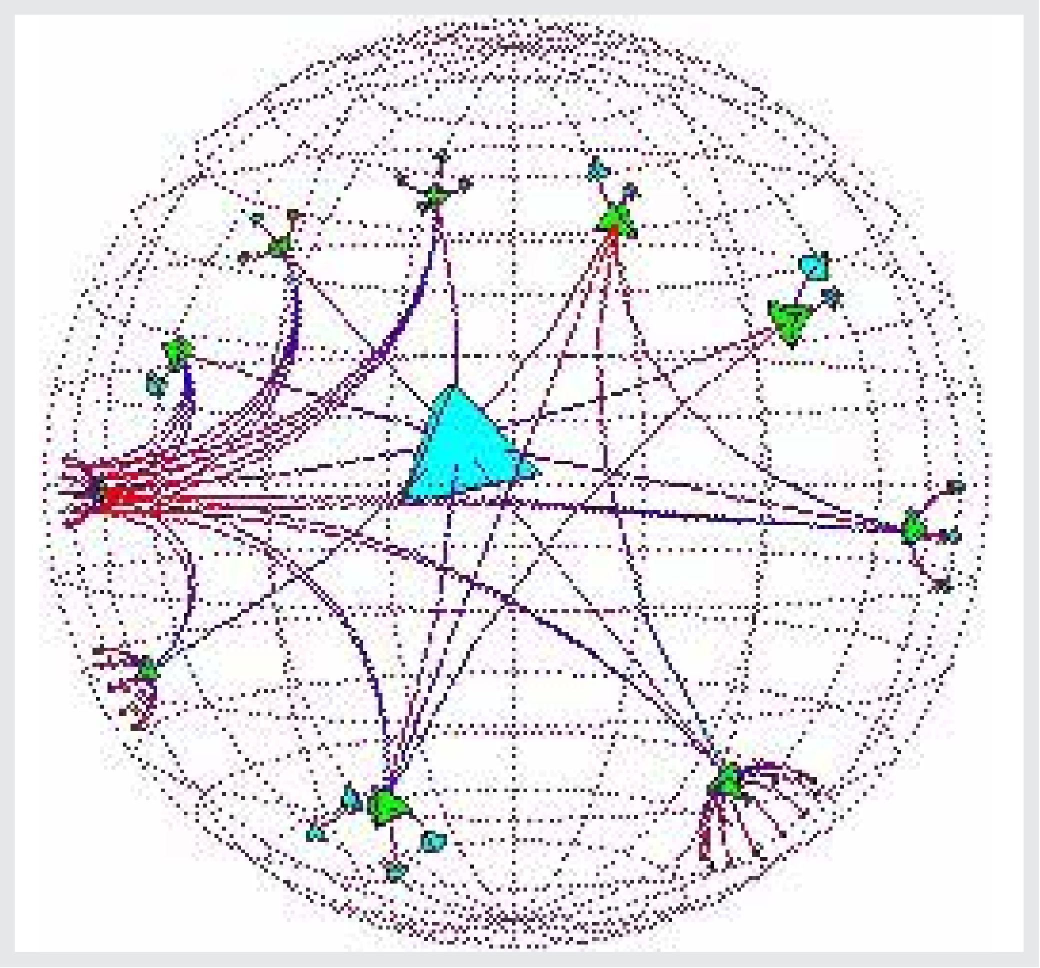
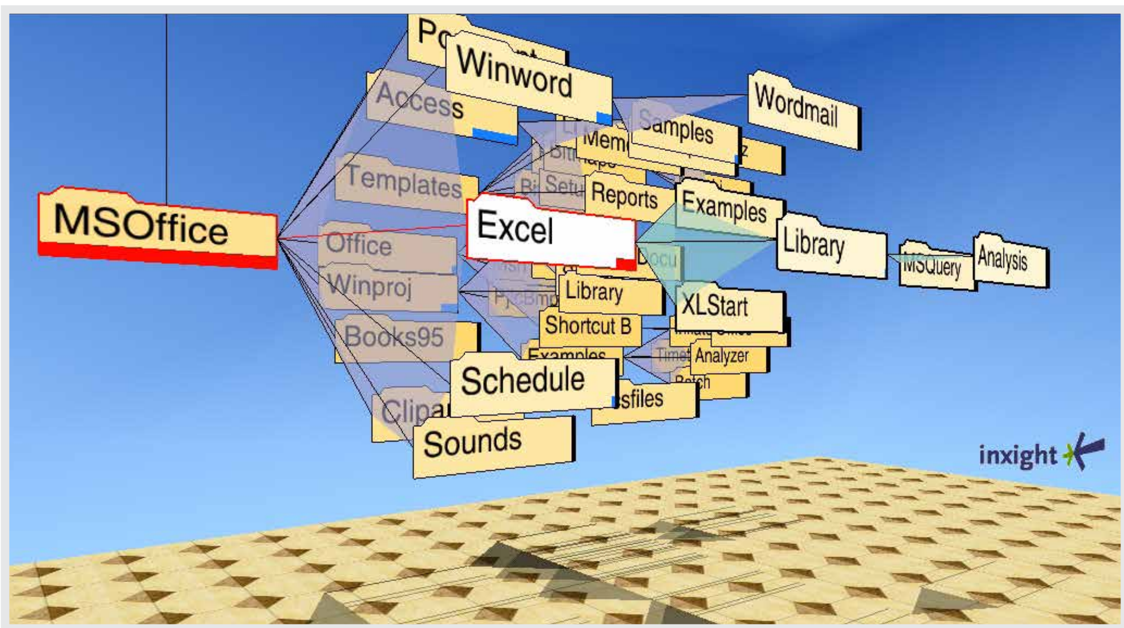




Figure 7.14 A-B-C-D: Hyperbolic visual transfer functions: (a) Hyperbolic visual transform function (Carpen-dale, 2001). (b) Hyperbolic tree (Lamping, Rao, and Pirolli, 1995) (c) 3D Hyperbolic (Munzner and Burchard, 1995; Munzner, 1998). (d) 3D Hyperbolic surface (Tallmina, 1997).

*Complex Visual Transfer Functions*. Some visual transfer functions are even more complex. Figure 7.15 shows a tree visualized in 3D as a cone tree (Robertson, Mackinlay, & Card, 1991), where each node has a hollow, 3D, rotatable circle of nodes beneath it. Figure 7.15.A. shows a small tree positioned obliquely, Figure 7.15.B. shows a much larger tree seen from the side. Touching an element in one of these trees will cause the circle holding the labels in that circle of the tree and all the circles above to rotate toward the user. The result is that the user will be able to read labels surrounding a point of interest, but natural perspective and occlusion will move into the background nodes of the tree more in the context. The visual transformation uses perspective as well as occlusion to attain a focus + context effect. The shift from focus to context is all done with geometric view transformations, but these are no longer described as a simple visual transfer of the sort in Figure 7.1.C.



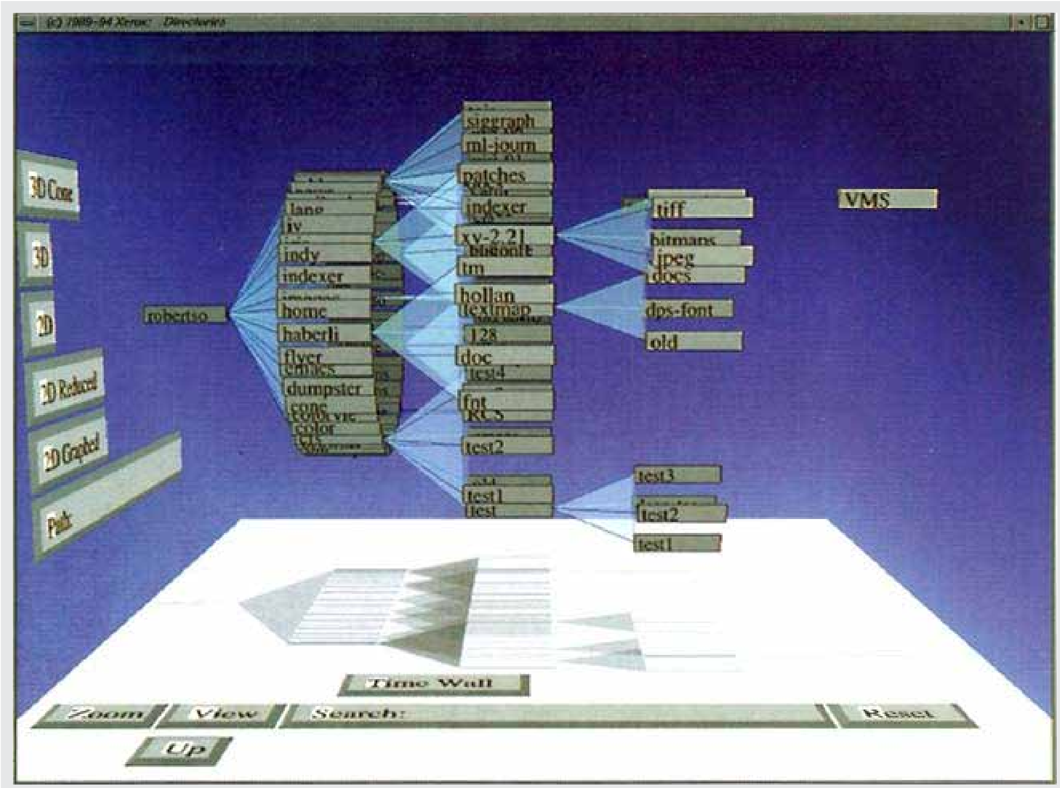





Figure 7.15 A-B: Cone tree: (b) Large cone tree from side. (a) Small cone tree showing perspective.

### 7.6.0.4 Degree-of-Interest Functions as Visual Mapping Transformations

By contrast with view transforms, visual mapping transforms use the content of data to generate physical form. Degree-of-Interest (DOI) functions assign an estimate of the momentary relevance to the user for each part of the data. This value is then used to modify the display dynamically. Suppose we have a tree of categories taken from Roget's Thesaurus, and we are interacting with one of these, "Hardness" (Figure 7.16.A). We calculate a degree-of-interest (DOI) for each item of the tree, given that the focus is on the node **Hardness**. To do this, we split the DOI into an intrinsic part and a part that varies with distance from the current center of interest and use a formula from Furnas (1982). Using a DOI function, the original tree can be collapsed to a much smaller tree (Figure 7.16.B) that preserves the focus and relevant parts of the context. How compact the resulting tree is depends on an interest threshold function. This could be a fixed number, but it could also be varied so that the resulting tree fits into a fixed size rectangle. In this way, DOI trees can be made to obtain the important user interface of property of spatial modularity. They can be assigned a certain size part of the screen resource and made to live within that space.

| | |
|---|---|
| Matter | Matter |
| ORGANIC vitality | ORGANIC vitality |
| Vitality in general | INORGANIC solid |
| Specific vitality | **Hardness** |
| Sensation in general | Softness |
| Specific sensation | Fluid |
| INORGANIC Solid | |
| **Hardness** | |
| Softness | |
| Fluid | |
| Fluid in general | |
| Specific fluid | |
| *(a) Categories from Roget's Thesaurus.* | *(b) Fisheye view of the categories when point of interest is centered on category* **Hardness**. |

FIGURE 7.16: Filtering with Degree-of-Interest function.

Of course, this is a small example for illustration. A tree representing a program listing, or a computer directory, or a taxonomy could easily have thousands of lines; a number that would vastly exceed what could fit on the display.

*DOI = Intrinsic DOI + Distance DOI*

Figure 7.17 shows schematically how to perform this computation for our example. The up-arrow indicates the presumed point of interest. We assume that the intrinsic DOI of a node is just its distance of the root (Figure 7.17.A). The distance part of the DOI is just the traversal distance to a node from the current focus node (Figure 7.17.B; it turns out to be convenient to use negative numbers for this computation, so that the maximum amount of interest is bounded, but not the minimum amount of interest. We add these two numbers together (Figure 7.17.C) to get the DOI of each node in the tree. Then we apply a minimum threshold of interest and only show nodes more interesting than that threshold. The result is the reduced tree in Figure 7.17.D. This is the sort of computation underlying Figure 7.16.D. The reduced tree gives local context around the focus node and progressively less detail farther away. But it does seem to give the important context.

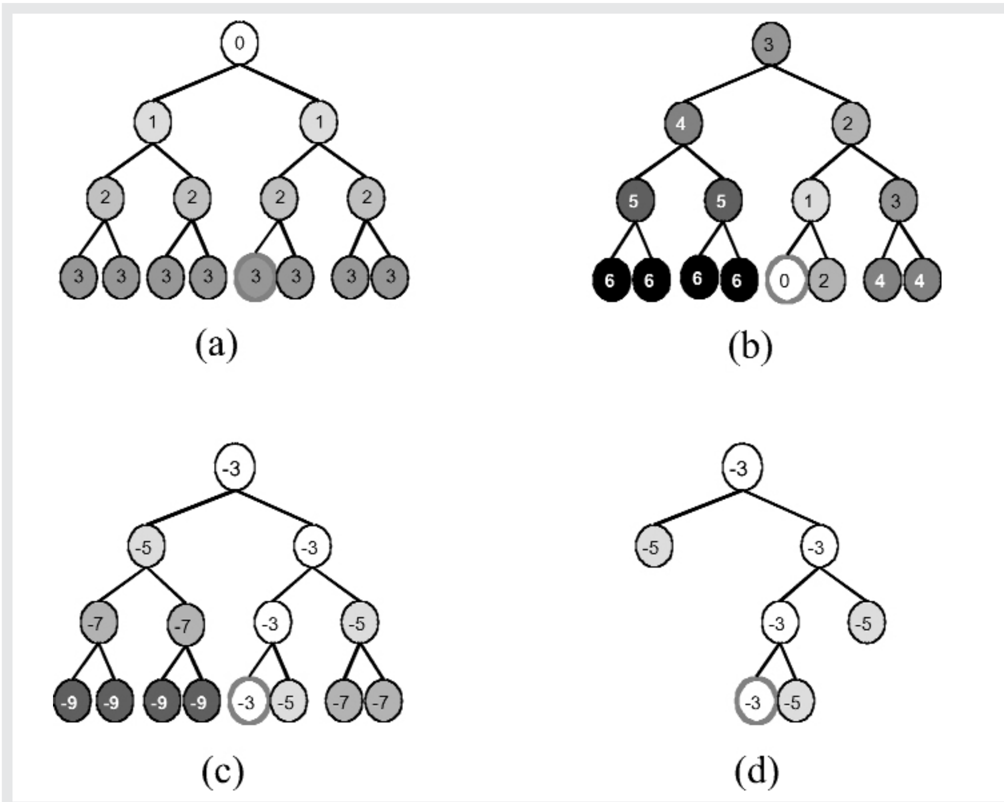



Figure 7.17: Computation of Degree-Of-Interest for a tree. (a) Intrinsic interest function. (b) Distance function. (c) Sum of (a) and (b). (d) Applying filtering function based on threshold to (c).



*Level-of-Detail Filtering with Degree-of-Interest Functions on multiple foci.* Figure 7.18 applies a version of these calculations to a tree with multiple focal points of interest comprising over 600,000 nodes. It is a demonstration that by blending a caching mechanism with the DOI calculation, calculations can be done on very large trees in a small fraction of a second, thereby allowing DOI trees to be used as a component of an animated interface to display contextualized, detail-filtered views of large datasets that will fit on the screen. If we assume the technique would work for at least a million nodes and that maybe 50 nodes would fit on the screen at one time, this demonstrates that we could get insightful, almost instantaneous views of trees 20,000 times larger than the screen would hold—a nice confirmation of the original bifocal display intuition.

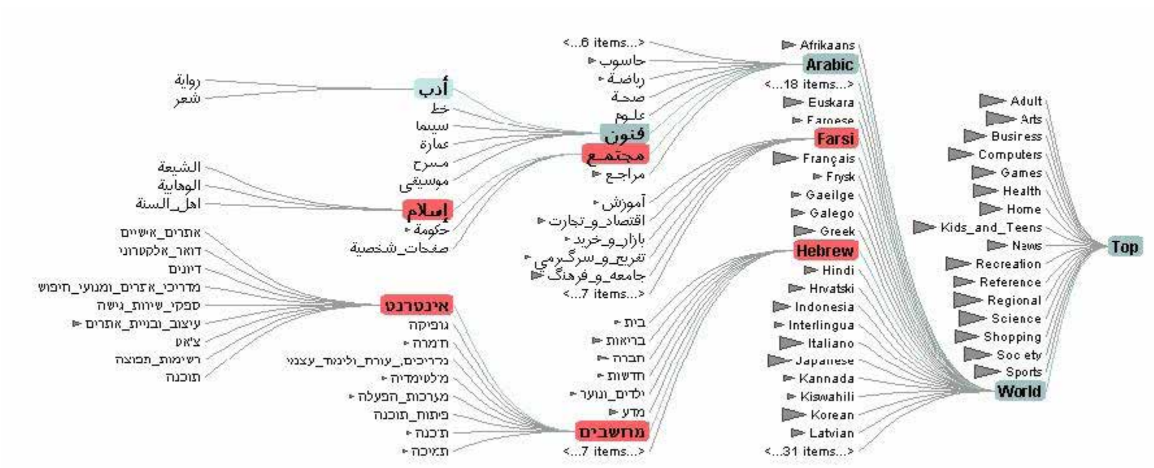



FIGURE 7.18: TreeBlock, a Degree-of-interest tree algorithm capable of computing and laying out very large trees at animation speeds. The tree here is shown with multiple foci on 600,000 nodes with mixed right-to-left and left-to-right text (Heer and Card, 2004).



*Re-Representation through semantic zooming and aggregation DOI functions.* Aside from level of detail filtering, it is possible to use the degree-of-interest information in many ways. In Figure 7.19, it is used (a) for level-of-detail filtering of nodes as previously discussed, (b) to size the nodes themselves, (c) to select how many attributes to display on a node, and (d) for semantic zooming. Semantic zooming substitutes smaller representation of about the same semantic meaning when the node is smaller. For example, the term "Manager" in Figure 7.19 might change to "Mgr." when the node is small.

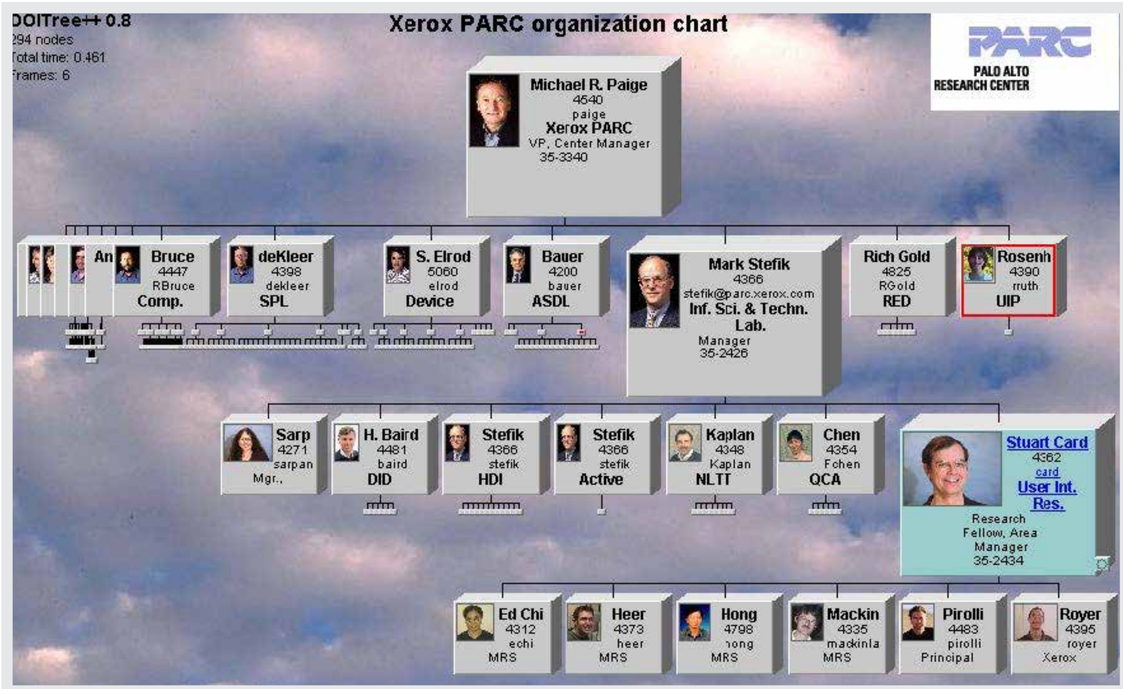
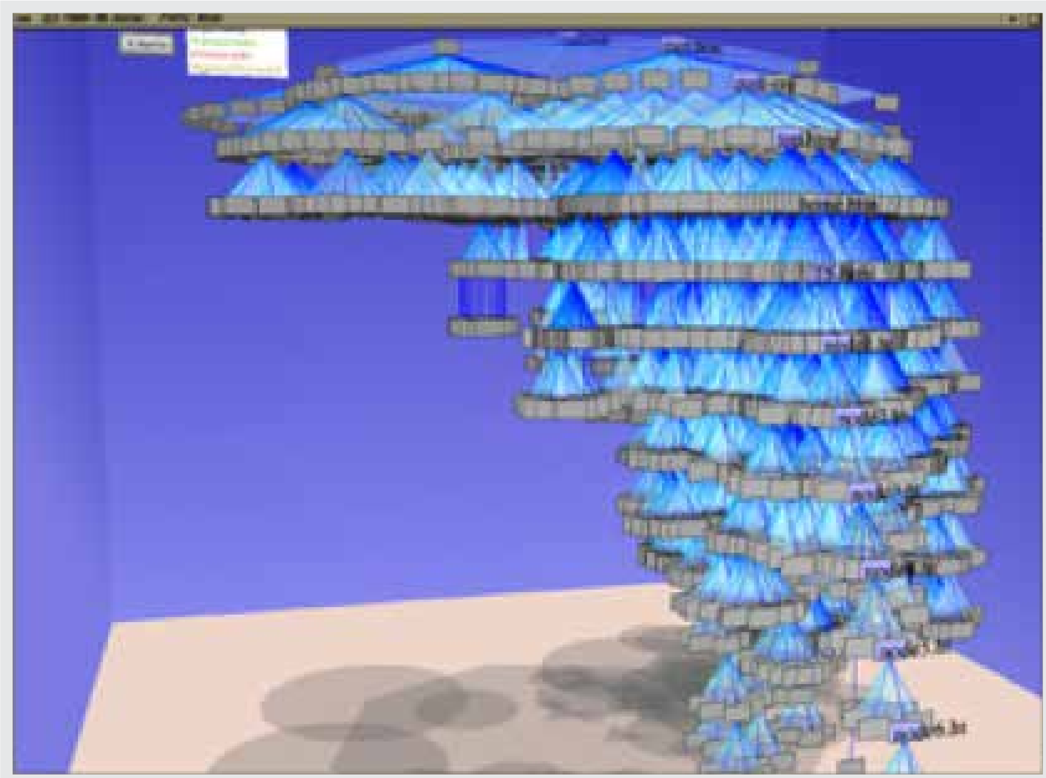
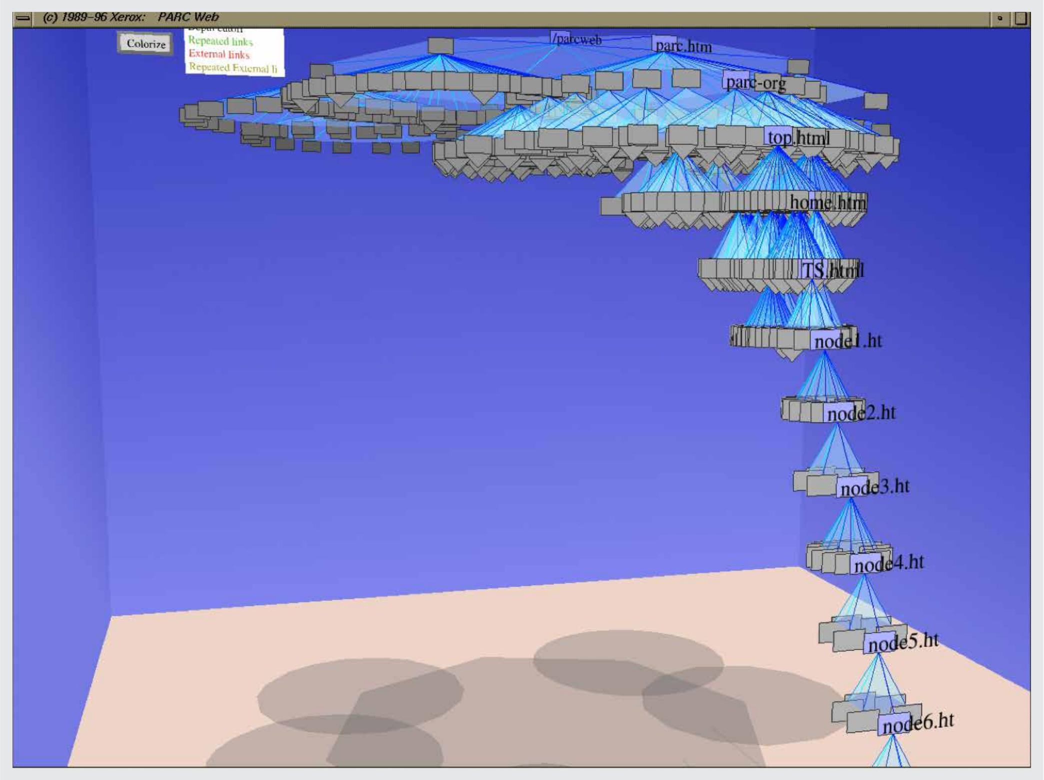



Figure 7.19: Degree-of-Interest calculation used to create an organization chart of PARC (in the early 2000's). Touching a box grows that box and boxes with whose degree of interest has been computed to increase in size and changes its contents; other boxes get smaller. Doing a search may result in multiple hits and cause several boxes to increase in size.



*Combining Visual Transformations with Degree-of-Interest Functions*. Of course both of the techniques we have been discussing can be combined. Figure 7.20 shows a cone tree containing all the files in Unix combined with a degree-of-interest function. The whole tree of files is shown in Figure 7.20.A. Selection focus on different files is shown in Figure 7.20.A and 7.20.B. Since Unix is a large system, this may be the first time anyone has ever "seen" *Unix*.




(c) 1989-96 Xerox: PARC Web
Colorize
Repeated links
External links
/parcweb
parc.htm
parc-org
top.html
home.htm
TS.html
node1.ht
node2.ht
node3.ht
node4.ht
node5.ht
node6.ht

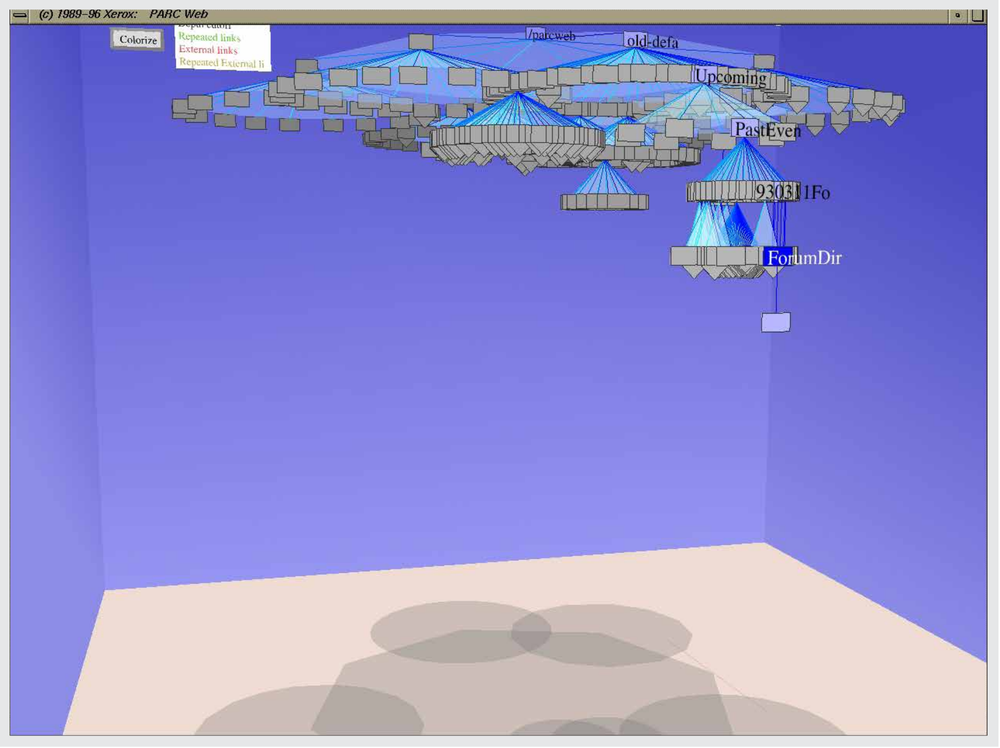





Figure 7.20: File in Unix visualized using a cone-tree combined with a degree-of-interest function. (a) Cone tree of all the files in Unix. (b) Use of degree-of-interest algorthm to select a subset of files. (c) Selection of another set of files.

### 7.6.0.5 The Current State of Focus + Context Displays

Focus + context techniques, inspired by the original work of Spence, Apperley, and Furnas, have turned out to be a rich source of ideas for dealing with information overload by pro-cessing local information in the context of global information structure. Spence and Apperley suggest some future direction for development. I agree with their suggestions and would like to suggest a few observations about what we have learned and what some of the op-portunities are. First the observations:

1. *Two reusable abstractions emerge for generating focus + context effects: (1) visual transfer functions and (2) degree-of-interest functions*. They structure much of the design space and help us generate new designs.

2. *But these principles may be interfered with by low-level vision phenomena*. For example, distortions of parallel lines may make the task more difficult. To compensate for this distortion, visual transfer functions can be given flat regions. Flat regions work, but may in turn give rise to an intermediate region between focal and context areas that creates a difficult-to-read area in the crucial near-focal region. For another example, the contextual part of the tree may form visual blobs, and the eye is attracted to visual blobs, leading it to spend time searching for things in the non-productive part of the tree (Pirolli, Card, & Van der Wege, 2003; Budiu, Pirolli, & Fleetwood, 2006). These uncontrolled effects may interfere with the task.We need to understand better low-level visual effects in focus + context displays.

3. *In general, we need to understand how focus + context displays provide cues to action or sensemaking in a task*. Distortion in a car rear view fisheye mirror is acceptable because the cue to action is the presence or absence or movement of some object in the mirrors field of view, indicating an unsafe situation. But if a fisheye display is used as part of a map viewer, the distorted bending of roads may not do well for cuing navigation. The difference is the task. Really we need to do a cognitive task analysis asking just what we are trying to get out of these displays and why we expect them to work. We have to understand better how focus + context displays work in the flow of the task.

4. *At large magnification ratios, focus + context displays work best when there is an emergent set of representations at the different aggregation levels*.

Using magnification alone can work for modest magnification levels. DOI filtering can work for large magnification ratios because its algorithm effectively shifts to a kind of higher-level aggregation. But the strength of focus + context displays is that they can tie together representations across aggregation levels.

5. Actually, these observations reflect a deeper set of issues. Focus + context displays trade on subtle interaction between the automatic, perceptually-oriented mechanisms of the user and the user's more effortful, cognitively-oriented mechanisms, sometimes called System 1 and System 2 (Kahneman, 2012) as well as on the subtle interaction of both of these systems with the demands of the task. The interaction of these mechanisms with the design of focus + context visualizations needs to be better understood. New opportunities for the development of these displays are: the integration with multi-touch input devices or multiple group displays or perhaps the use in automobiles or medical operating rooms. Focus + context displays are about the dynamic partitioning of bandwidth and attention. New information streams for problems, and new input devices for control should insure that this is still a fertile area.

### 7.6.0.6 References

1. Budiu, Raluca, Pirolli, Peter, & Fleetwood, Michael (2006). Navigation in degree of interest trees. *AVI 2006, Proceedings of the working conference on Advanced Visual Interfaces,* 457-462. New York: ACM.
2. Card, Stuart K., Hong, Lichen, Mackinlay, Jock D. (2004). 3Book: A 3D electronic smart Book. *AVI 2004, Proceedings of the conference on Advanced Visual Interfaces:* 303-307.

3. Card, S. K., Mackinlay, J. D., & Shneiderman, B. (1999). *Readings in Information Visualization*. San Francisco, CA: Morgan-Kaufmann.

4. Card, Stuart K. and Nation, David (2002). Degree-Of-Interest Trees: A component of an attention-reactive user interface. *AVI 2002, Proceeding of the Conference on Advanced Visual Interfaces (Trento, Italy, May 22-24, 2002)*, 231-245.

5. Card, S. K., Robertson, G. G., and York, W. (1996). The WebBook and the Web Forager. An information workspace for the World-Wide Web. *Proceedings of CHI '96, ACM Conference on Human Factors in Software*, New York: ACM, 111–117.

6. Carpendale, M.S.T. and Montagnese, Catherine (2001). Rubber sheet visual transform functions. *In UIST 2001, Proceedings of the ACM Symposium on User Interface Software and Technology (Orlando, FL)*. New York: ACM, 61-70.

7. Carpendale (2006/2012, December 6). Elastic Presentation. *Sheelagh Carpenter*. Retrieved June 24, 2012 from http://pages.cpsc.ucalgary.ca/~sheelagh/wiki/pmwiki.php?n=Main.Presentation .

8. Furnas, G. (1986). Generalized fisheye views. *CHI '86, Proceeding of the Conference on Human Factors in Computing Systems (Boston)*. New York: ACM, 16-23.

9. Furnas, G. W. (1982). The FISHEYE view: A new look at structured files. Technical Memorandum #82-11221-22, Bell Laboratories, Oct. 18.

10. Furnas, G. W. (1992/1999). The FISHEYE view: A new look at structured files. In Stuart. K. Card, Jock D. Mackinlay, & Ben Shneiderman, *Readings in Information Visualization* (pp. 312-330). San Francisco, CA: Morgan-Kaufmann.

11. Heer, Jeffrey and Card, Stuart K. (2003). Information visualization & navigation: Efficient user interest estimation in fisheye views. *CHI '03 Extended Abstracts on Human Factors in Computing Systems,* 836-837.

12. Heer, Jeffrey & Card, Stuart K. (2004). DOITrees Revisited: Scalable, space-constrained visualization of hierarchical data. AVI 2004. Proceeding of the Conference on Advanced Visual Interfaces (Trento, Italy),

13. Kahneman, Daniel (2011). *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux.

14. Lamping, J., Rao, R. and Pirolli, P. A focus + context technique based on hyperbolic geometry for visualizing large hierarchies. *CHI '95 P*roceedings of the SIGCHI Conference on Human factors in Computing Systems.

15. Leung, Y. W. and Apperley, Mark (1994). A Review and Taxonomy of Distortion-Oriented Presentation Techniques. In *ACM Transactions on Computer-Human Interaction, 1(2): 126-160*.

16. Mackinlay, Jock D., Robertson, George G., & Card, Stuart K. (1991). The perspective wall: Detail and context smoothly. In *CHI '91, ACM Conference on Human Factors in Computing Systems*. New York: ACM, 173–179.

17. Munzner, Tamara, and Burchard, Paul (1995). Visualizing the structure of the World Wide Web in 3D hyperbolic space. *Proceedings of VRML '95, (San Diego, California, December 14-15)* and special issue of *Computer Graphics*, New York: ACM SIGGRAPH, pp. 33-38.

18. Munzner, Tamara (1998). Exploring large graphs in 3D hyperspace. IEEE Computer Graphics and Applications 18(4):18-23.

19. Pirolli, Peter, Card, Stuart K., a& Van Der Wege, Mija M. (2003). The effects of information scent on visual search in the hyperbolic tree browser. *ACM Transactions on Computer-Human Interaction (TOCHI): 10*(1). New York: ACM.

20. Resnikoff, H. L. (1987). *The Illusion of Reality*. New York: Springer-Verlag.

21. Robertson, George G., Mackinlay, Jock D. (1993). The document lens. In *ACM UIST '93, Proceedings of the 6th Annual ACM Symposium on User Interface Software and Technology*. New York: ACM, 101-108.

22. Robertson, George G., Mackinlay, Jock D., & Card, Stuart K. (1991). Cone trees: Animated 3D visualizations of hierarchical information. *CHI '91 ACM Conference on Human Factors in Computing Systems,* 189–194. New York: ACM.

23. Sarkar, M. and Brown, M.H. (1994). Graphical fisheye views*, CACM* 37(12): 73-84.

24. Spence, R. (1982). The Bifocal Display. Video, Imperial College London.

25. Tallmina, Daina (1997). Crochet model of hyperbolic plane. Fig, The Institute for Figuring. Taken from web page http://www.theiff.org/oexhibits/oe1e.html on June 14, 2012.

## 7.7 COMMENTARY BY LARS ERIK HOLMQUIST

How to cite this commentary in your report

### Lars Erik Holmquist

© *Lars Erik Holmquist*

Lars Erik Holmquist is Professor in Media Technology at Södertörn University, manager of the Interaction Design and Innovation lab at the Swedish Institute of Computer Science, and a Research Leader at the Mobile Life VINN Excellence Centre in Kista, Sweden. He previously led research groups at the Viktoria Institute and the Interactive Institute. He received his M.Sc. in Computer Scienc...

Lars Erik Holmquist
Lars Erik Holmquist is a member of The Interaction Design Foundation

When revisiting the original videos by Spence and Apperley, it is remarkable how fresh and practical their ideas still are - and this goes for not just the principles of the Bifocal display itself, but also the human-computer interaction environment that they envisioned. A few years ago I organized a conference screening of classic research videos, including Spence and Apperley's envisionment of a future *Office of the Professional*. For entertainment purposes, the screening was followed by Steven Spielberg's science fiction movie MINORITY REPORT. In the fictional film, we could see how the hero (played by Tom Cruise) interacted with information in a way that seemed far beyond the desktop computers we have today - but in many ways very similar to Spence and Apperley's vision of the future office. So ahead of their time were these researchers that when these works were shown in tandem, it became immediately obvious how many of the ideas in the 1981 film were directly reflected in a flashy Hollywood vision of the future - created over 20 years later!

It is hard for us to imagine now, but there was a time when the desktop computing paradigm, also called Windows-Icons-Mouse-Pointers or WIMP, was just one of many competing ideas for how we would best interact with digital data in the future. Rather than pointing and clicking with a disjointed, once-removed device like the mouse, Spence and Apperley imagined interactions that are more in line with how we interact with real-world objects - pointing directly at them, touching them on the screen, issuing natural verbal commands. Of the many ideas they explored, the general theme was interaction with large amounts information in ways that are more natural than viewing it on a regular computer screen - something they likened to peeking through a small window, revealing only a tiny part of a vast amount of underlying data.

The Bifocal display is based on some very simple but powerful principles. By observing how people handle large amounts of data in the real, physical world, the inventors came up with a solution for mitigating the same problem in the virtual domain. In this particular case, they drew upon an observation of human vision system - how we can keep many things in the periphery of our attention, while having a few in the focus - and implemented this electronically. They also used a simple optical phenomenon, that of perspective; things in the distance are smaller than those that are near. Later, other physical properties have also been applied to achieve a similar effect, for instance the idea of a "rubber sheet" that stretches and adapts to an outside force, or that of a camera lens that creates a "fisheye" view of a scene (e.g. Sarkar and Brown 1994).

All of these techniques can be grouped under the general term of *focus+context visualizations*. These visualizations have the potential to make large amounts of data comprehensible on computers screens, which are by their nature limited in how much data they can present, due to factors of both size and resolution. However, powerful as they may be, there are also some inherent problems in many of these techniques. The original Bifocal display assumes that the material under view is arranged in a 1-dimensional layout, which can be unsuitable for many important data sets, such as maps and images. Other fisheye and rubber sheet techniques extended the principles to 2-dimensional data, but still require an arrangement based on fixed spatial relationships rather than more logically based ones, such as graphs. This has been addressed in later visualization techniques, which allow the individual elements of a data set (e.g. nodes in a graph) to move more freely in 2-dimensional space while keeping their logical arrangement (e.g. Lamping et al 1995).

Furthermore, for these techniques to work, it is necessary to assume that the material outside the focus is not overly sensitive to distortion shrinking, or that it at least can be legible even when some distortion is applied. This is not always true; for instance, text can become unreadable if subjected to too much distortion and/or shrinking. In these cases, it may be necessary to apply some other method than the purely visual to reduce the size of the material outside the focus. One example of how this can be done is *semantic zooming,* which can be derived from the *Degree of Interest* function in Furnas' generalized fisheye views (Frunas 1986). With semantic zooming, rather than graphically shrinking or distorting the material outside the focus, important semantic features are extracted and displayed. A typical application would be to display the headline of a newspaper article rather than a thumbnail view of the whole text. Semantic zooming is now common in maps, where more detail - such as place names and small roads - gradually gets revealed as the user zooms in.

There have been many approaches that try to mitigate these problems. In my own work, using a similar starting point to Spence and Apperley and also inspired by work by Furnas, Card and many others, I imagined a desk covered with important papers. One or two would be in the center of attention as they were being worked on; the rest would be spread around. However, unlike other bifocal displays they would not form a continuous display, but be made up of discrete objects. On a computer screen, the analog would be to have one object in the middle in readable size, and the others shrunk to smaller size arranged on the surrounding area. By arranging the individual pages in a left-to-right, top-to-bottom fashion it became possible to present a longer text, such as a newspaper article or a book (see figure 1). The user could then click on a relevant page to bring it into focus, or use the keyboard to flip through the pages (Figure 2). This technique was called *Flip Zooming,* as it mimicked flipping the pages in a book. The initial application was a Java application for web browsing, called the *Zoom Browser* (Holmquist 1997). Later we worked to adapt the same principle to smaller displays, such as handheld computers. Because the screen real-estate on these devices was even smaller, just shrinking the pages outside the focus was not feasible - they would become too small to read. Instead, we applied computational linguistics principles to extract only the most important important keywords of each section, and present these to give the viewer an overview of the material. This was implemented as a web browser for small terminals, and was one of the first examples of how to handle large amounts of data on such devices (Björk et al. 1999).

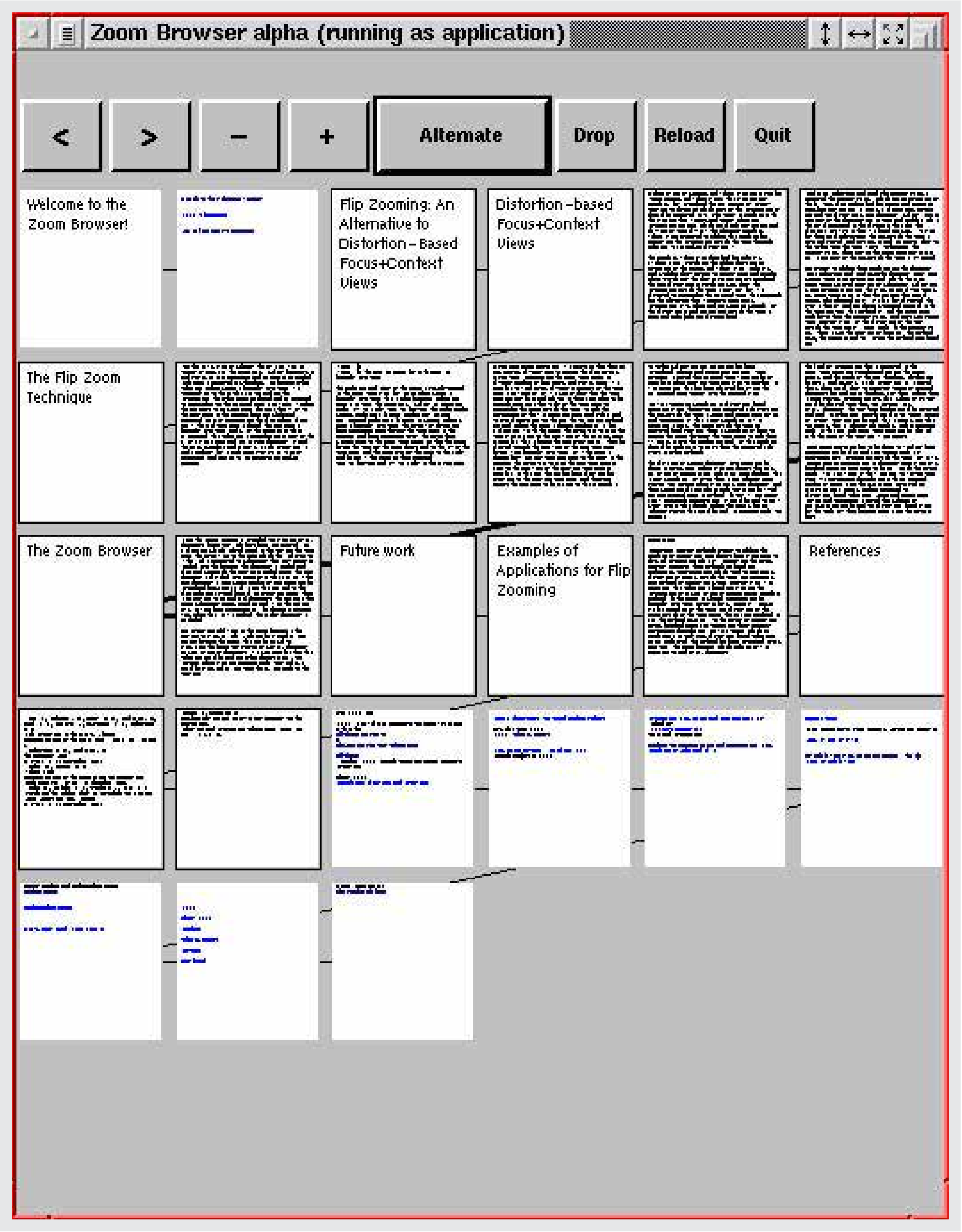



FIGURE 7.1: Flip zooming view of a large document, with no page zoomed in.

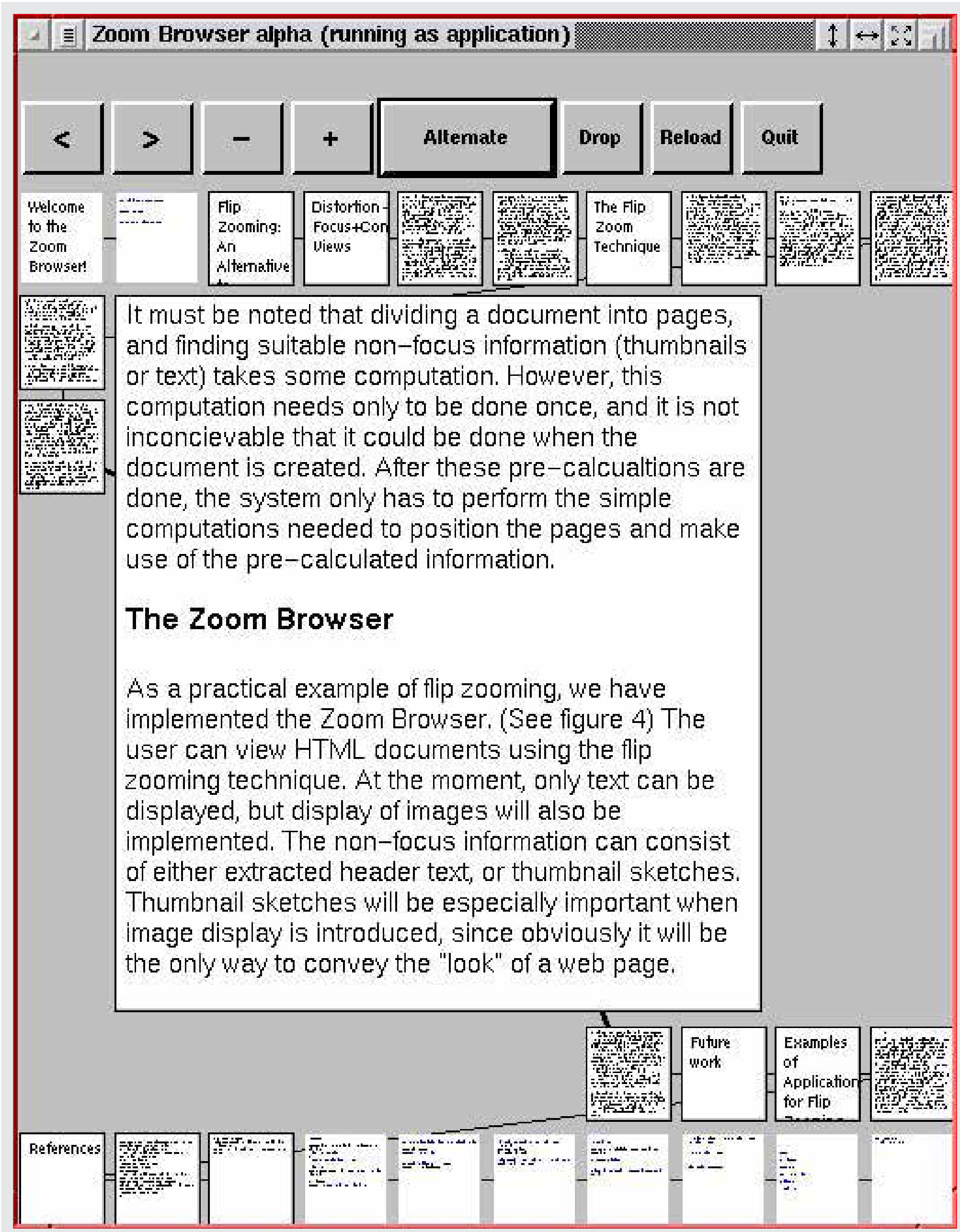



Figure 7.2: Flip zooming with a page zoomed in. Note the lines between pages to denote order!.

Another problem with visualizing large amounts of data, is that of size versus resolution. Even a very large display, such as a projector or big-screen plasma screen, will have roughly the same number of pixels as a regular computer terminal. This means that although we can blow up a focus+context display to wall size, the display might not have enough detail to properly show the important information in the focus, such as text. Several projects have attempted to combine displays of different sizes resolutions in order to show both detail and context at the same time. For instance, the *Focus Plus Context Screen* positioned a high-resolution screen in the centre of a large, projected display (Baudisch et al 2005). This system made it possible to provide low-resolution overview of a large image, e.g. a map, with a region of higher resolution in the middle; the user could then scroll the image to find the area of interest. A similar approach was found in the *Ubiquitous Graphics* project,where we combined position-aware handheld displays with a large projected display. Rather than scrolling an image around a statically positioned display, users could move the high-resolution display as a window or “magic lens” to show detail on an arbitrary part of the large screen (see Figure 3). These and several other projects point to a device ecology where multiple screens act in tandem as input/output devices. This would allow for collaborative work in a much more natural style than allowed for by the single-user desktop workstations, in a way that reminds us of the original Spence and Apperley vision.

FIGURE 7.3: The ubiquitous graphics system provided a freely movable high-resolution display, that acted as an interactive "magic lens" to reveal detailed information anywhere on the larger display.



After over 20 years of WIMP desktop computing, the Bifocal display and the ideas derived from it are therefore in many ways more relevant than ever. We live in a world where multiple displays of different resolutions and sizes live side by side, much like in Spence and Apperley's vision of the future office. New interaction models have opened up new possibilities for zooming and focus+context based displays. For instance, multitouch devices such as smartphones and tablets make it completely intuitive to drag and stretch a virtual "rubber sheet" directly on the screen, instead of the single-point, once-removed interaction style of a mouse.

I believe that this new crop of devices presents remarkable opportunities to revisit and build upon the original visualization ideas presented in Spence's text, and that we may have only seen the very start of their use in real-world applications.

## References


- Björk, S., Holmquist, L.E., Redström, J., Bretan, I., Danielsson, R., Karlgren, J. and Franzén, K. *WEST: A Web Browser for Small Terminals*. Proc. ACM Conference on User Interface Software and Technology (UIST) '99, ACM Press, 1999.
- Baudisch, P., Good, N., and Stewart, P. *Focus Plus Context Screens: Combining Display Technology with Visualization Techniques*. In Proceedings of UIST 2001, Orlando, FL, November 2001, pp.31-40.
- Furnas, G.W. *Generalized fisheye views*. CHI '86 Proceedings of the SIGCHI conference on Human factors in computing systems.
- Holmquist, L.E. *Focus+context visualization with flip zooming and the zoom browser*. CHI '97 extended abstracts on Human factors in computing systems.
- Lamping, J., Rao, R. and Pirolli, P. *A focus+context technique based on hyperbolic geometry for visualizing large hierarchies*. CHI '95 Proceedings of the SIGCHI conference on Human factors in computing systems.
- Sanneblad, J. and Holmquist, L.E. *Ubiquitous graphics: combining hand-held and wall-size displays to interact with large images*. AVI '06 Proceedings of the working conference on Advanced visual interfaces.
- Sarkar, M. and Brown, M.H. *Graphical Fisheye Views*. Commun. ACM 37(12): 73-84 (1994)

## 7.5 BEHIND THE SCENES

*Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.*


*Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.*


Lift
British Rail
Concourse and
street level
Ticket office opening hours




Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.


Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.

*Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.*

*Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.*

*Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.*

## 7.8 REFERENCES

*Apperley, Mark and Leung, Y. K. (1993b): A taxonomy of distortion-oriented techniques for data presentation. In: Salvendy, Gavriel and Smith, M. J. (eds.). "Advances in Human Factors/Ergonomics Vol 19B, Human-Computer Interaction: Software and Hardware Interfaces". Amsterdam, Holland: Elsevier Science Publisherspp. 105-109*

*Apperley, Mark and Leung, Y. K. (1993a). A Unified Theory of Distortion-Oriented Presentation Techniques. Massey University*

*Apperley, Mark and Spence, Robert (1980). Video: bifocal display concept video. Retrieved 6 June 2013 from http://www.interaction-design.org/tv/bifocal_display_concept_video.html*

*Apperley, Mark and Spence, Robert (1981): A Professional's Interface Using the Bifocal Display. In: Proceedings of the 1981 Office Automation Conference 1981. pp. 313-315*

*Apperley, Mark, Spence, Robert and Wittenburg, Kent (2001): Selecting One from Many: The Development of a Scalable Visualization Tool. In: HCC 2001 - IEEE CS International Symposium on Human-Centric Computing Languages and Environments September 5-7, 2001, Stresa, Italy. pp. 366-372*

*Apperley, Mark, Tzavaras, I. and Spence, Robert (1982): A Bifocal Display Technique for Data Presentation. In: Eurographics 82 Proceedings 1982, Amsterdam. pp. 27-43*

*Bederson, Benjamin B., Clamage, Aaron, Czerwinski, Mary and Robertson, George G. (2004): DateLens: A fisheye calendar interface for PDAs. In ACM Transactions on Computer-Human Interaction, 11 (1) pp. 90-119*

*Buxton, Bill (2007): Sketching User Experiences: Getting the Design Right and the Right Design. Morgan Kaufmann*

*Cohen, Marcelo, Brodlie, Ken and Phillips, Nick (): Hardware-accelerated distortion for volume visualisation in medicine. In: Proceedings of the 4th IEEE EMBSS UKRI PG Conference on Biomedical Engineering and Medical Physics 2005 . pp. 29-30*

*Farrand, William A. (1973). Information display in interactive design, Doctoral Thesis. University of California at Los Angeles*

*Forlines, Clifton and Shen, Chia (2005): DTLens: multi-user tabletop spatial data exploration. In: Proceedings of the 2005 ACM Symposium on User Interface Software and Technology 2005. pp. 119-122*

*Furnas, George W. (1986): Generalized Fisheye Views. In: Mantei, Marilyn and Orbeton, Peter (eds.) Proceedings of the ACM CHI 86 Human Factors in Computing Systems Conference April 13-17, 1986, Boston, Massachusetts. pp. 16-23*

*Guiard, Yves and Beaudouin-Lafon, Michel (2004): Target acquisition in multiscale electronic worlds. In International Journal of Human-Computer Studies, 61 (6) pp. 875-905*

*Lamping, John and Rao, Ramana (1994): Laying Out and Visualizing Large Trees Using a Hyperbolic Space. In: Szekely, Pedro (ed.) Proceedings of the 7th annual ACM symposium on User interface software and technology November 02 - 04, 1994, Marina del Rey, California, United States. pp. 13-14*

*Leung, Ying K. and Apperley, Mark (1993): E{cubed}: Towards the Metrication of Graphical Presentation Techniques for Large Data Sets. In: East-West International Conference on Human-Computer Interaction: Proceedings of the EWHCI93 1993. pp. 9-26*

*Leung, Ying K. and Apperley, Mark (1993): Extending the Perspective Wall. In: Proceedings of OZCHI93, the CHISIG Annual Conference on Human-Computer Interaction 1993. pp. 110-120*

*Leung, Y. W. and Apperley, Mark (1994): A Review and Taxonomy of Distortion-Oriented Presentation Techniques. In ACM Transactions on Computer-Human Interaction, 1 (2) pp. 126-160*

*Leung, Ying K., Spence, Robert and Apperley, Mark (1995): Applying Bifocal Displays to Topological Maps. In International Journal of Human-Computer Interaction, 7 (1) pp. 79-98*

*Mackinlay, Jock D., Robertson, George G. and Card, Stuart K. (1991): The Perspective Wall: Detail and Context Smoothly Integrated. In: Robertson, Scott P., Olson, Gary M. and Olson, Judith S. (eds.) Proceedings of the ACM CHI 91 Human Factors in Computing Systems Conference April 28 - June 5, 1991, New Orleans, Louisiana. pp. 173-179*

*Mazza, Riccardo (2009): Introduction to Information Visualization. Springer*

*Modine, Austin (2008). Apple patents OS X Dock. Retrieved 9 November 2010 from The Register: http://www.theregister.co.uk/2008/10/08/apple_patents_osx_dock/*

*Rao, Ramana and Card, Stuart K. (1994): The Table Lens: Merging Graphical and Symbolic Representations in an Interactive Focus+Context Visualization for Tabular Information. In: Adelson, Beth, Dumais, Susan and Olson, Judith S. (eds.) Proceedings of the ACM CHI 94 Human Factors in Computing Systems Conference April 24-28, 1994, Boston, Massachusetts. pp. 318-322*

*Sarkar, Manojit, Snibbe, Scott S., Tversky, Oren J. and Reiss, Steven P. (1993): Stretching the Rubber Sheet: A Metophor for Visualizing Large Layouts on Small Screens. In: Hudson, Scott E., Pausch, Randy, Zanden, Brad Vander and Foley, James D. (eds.) Proceedings of the 6th annual ACM symposium on User interface software and technology 1993, Atlanta, Georgia, United States. pp. 81-91*

*Spence, Robert (2007): Information Visualization: Design for Interaction (2nd Edition). Prentice Hall*

*Spence, Robert (2001): Information Visualization. Addison Wesley*

*Spence, Robert and Apperley, Mark (1982): Data Base Navigation: An Office Environment for the Professional. In Behaviour and Information Technology, 1 (1) pp. 43-54*

*Tidwell, Jenifer (2005): Designing Interfaces: Patterns for Effective Interaction Design. O'Reilly and Associates*

*Tobler, W. R. (1973): A continuous transformation useful for districting. In Annals of the New York Academy of Sciences, 219 p. 215–220*

*Ware, Colin (2004): Information Visualization: Perception for Design, 2nd Ed. San Francisco, Morgan Kaufman*

## About the authors

### Robert Spence

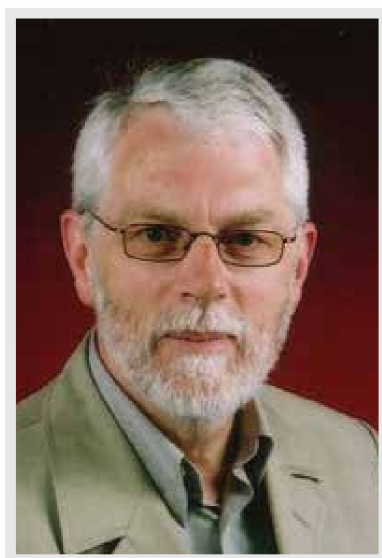
© *Robert Spence*

HAS ALSO PUBLISHED UNDER THE NAME OF:
"Bob Spence"

Personal Homepage:
http://www.ee.ic.ac.uk/r.spence/

Bob Spence is Professor Emeritus of Information Engineering at Imperial College London. Bob Spence's research has ranged from engineering design to human-computer interaction,and often with the manner in which the latter can enhance the former. Notable contributions, usually in collaboration with colleagues, include the powerful generalized form of Tellegen's Theorem; algorithms for improving the manufacturing yield of mass-produced circuits; and, in the field of Human-computer Interaction, the invention of the first focus+context technique, the Bifocal Display (aka Fisheye lens). The novel Attribute and Influence Explorers provide examples of novel information visualization tools that have wide application, including engineering design. Interactive computer graphics allows the electronic circuit designer to sketch the familiar circuit diagram on a computer display. This potential was pioneered by Bob and his colleagues in the late 1960s and eventually, in 1985, led to the commercially available MINNIE system developed and marketed by a company of which Bob was chairman and a founding director. More recently, Bob's research has focused on the topic of Rapid Serial Visual Presentation in which a collection of images is presented sequentially and rapidly to a user who may be searching for a particular image. This activity is similar to the riffling of a book's pages.

## Mark Apperley



Has also published under the name of:
"Mark D. Apperley"

Personal Homepage:
cs.waikato.ac.nz/genquery.php?linklevel=4&linklist=CS&linkname=A_to_I&linktype=report&listby=Name&foldername=A_to_I&reportdb=People&report=People&where=Department%20=%20'Computer%20Science'%20&&%20Type%20!=%20'Postgraduate'%20&&%20(Last_Name%20

Current place of employment:
University of Waikato

Mark Apperley has been working in the field of HCI for more than 30 years. In the 1970's he worked on the MINNIE interactive CACD system with Bob Spence, pioneering a range of interaction and information visualisation techniques, including dynamic exploration and percent done indicators. Also with Bob Spence he devised the bifocal display (1980) and the Lean Cuisine notation for menu description (1988). More recently his research has focussed on systems supporting collaborative work, and on techniques for large screen interaction. Mark is Professor of Computer Science and Dean of the School of Computing and Mathematical Sciences at the University of Waikato, Hamilton, New Zealand.

## Your notes and thoughts on chapter 7

Record your notes and thoughts on this chapter. If you want to share these thoughts with others online, go to the bottom of the page at:
http://www.interaction-design.org/encyclopedia/bifocal_display.html

**NOTES**

# Chapter 8

# Contextual Design

*by Karen Holtzblatt and Hugh R. Beyer.*

*Contextual Design* is a structured, well-defined user-centered design process that provides methods to collect data about users in the field, interpret and consolidate that data in a structured way, use the data to create and prototype product and service concepts, and iteratively test and refine those concepts with users. This is the core of the Contextual Design philosophy - understand users in order to find out their fundamental intents, desires, and drivers. But these are invisible to the users - so the only way to glean them is to go out in the field and talk with people

Although based on theories from several disciplines, including anthropology, psychology and design, Contextual Design was designed for practical application with commercial design teams.

Since its original development, Contextual Design has been applied in a variety of industries and also used as a vehicle to teach user-centered design principles in engineering and design programs.

Contextual Design has primarily been used for the design of computer information and IT systems, including hardware (Curtis et al 1999) and software (Rockwell 1999). Parts of Contextual Design have been adapted for use as a field usability evaluation method (McDonald et al 2006). Contextual Design has also been applied to the design of digital libraries and other learning technologies (Notess 2005, Notess 2004). Contextual Design has also been used in a variety of other industries, including web applications, process reengineering, consumer product design, manufacturing, and automotive and medical device design, to name just a few.

Contextual design has also been widely used as a means of teaching user-centered design and human-computer interaction at the university level (Weinberg and Stephen 2002, Larusdottir 2006).

## 8.1 MOTIVATIONS AND KEY PRINCIPLES

A small number of key principles shaped the development of Contextual Design and provide the key motivations for its use as a design tool.

### 8.1.1 Principle: System design must support and extend users' work practice

Contextual Design is rooted in the observation that any technology or system is always situated in a larger environmental context - and that introduction of new solutions invariably changes the environment for its users. In Contextual Design, the term *work practice* refers to the complex and detailed set of behaviors, attitudes, goals and intents that characterize a set of users in a particular environment. All manner of activities and design domains are characterized by work practice - not only workplaces. For example, there are obviously work practices associated with business pursuits like office work, but there are also "work practices" associated with life events such as making purchases as a consumer, driving an automobile, playing music and even watching television. A central tenet of Contextual Design is that *any technology, product or system must be designed to support and extend its users' work practice*. If it does so well, it will be accepted and valued; if it fails to do so, it will cause dissatisfaction, frustration, avoidance and workarounds

Implications for the designer: To create a successful product, first be aware of users' work practice and design for it explicitly.

### 8.1.2 Principle: People are experts at what they do - but are unable to articulate their own work practice

Complicating the designer's job are two facts about work practice. The first is that *people are not consciously aware of their own work practice*; all of their knowledge is tacit. This is especially true when people are taken out of the context of their everyday environment. It is only when users are immersed in normal contexts of use that they can become aware of their own work practice - what they do in detail and why. They become "aware in the doing," as Michael Polanyi puts it (Polanyi 1958).

The second is that *work practice is complex and varied*, and that *useful design data are hidden in everyday details*. Many systems fall short of expectations because they fail to take into considerations seemingly insignificant details of work practice - details that are not consciously available to users when they are not engaged in the ongoing work.

Contextual Design holds that design team members must go into the field and observe and talk with users in their natural work or life environments - their natural contexts - in order to understand work practice. This is the principle of *context* from which the process draws its name. This aspect of Contextual Design leverages the work of earlier ethnographic methodologies (Garfinkel 1967) but extends it in important ways.

Implications for the designer: Use field interviews to reveal tacit aspects of users' work practice - the motivations, workarounds, and strategies that they may never articulate, but structure their work.

### 8.1.3 Principle: Good design requires partnership and participation with users

Even while in context, users are not always able to intuit and articulate their own behaviors and detailed motivations. And so Contextual Design prescribes interviews that are not pure ethnographic observations, but involve the user in discussion and reflection on their own actions, intents, and values.

The interviewer actively questions the user and *partners* with them to draw out and understand their work practice in detail. The interviewer thus does not enter with a preformed list of questions, as in a survey or focus group, but rather adopts a *master-apprentice relationship model*, seeking to understand the user's work as an apprentice would from a master, as the work is ongoing.

This key concept of partnership also comes into play in Contextual Design's use of paper prototypes and short iterations with users to work out detailed design. The thinking behind Contextual Design's iterative prototyping evolved in conjunction with, and influenced, the development of participatory design techniques in the 1980's and 1990's (Schuler and Namioka 1993).

Implications for the designer: Don't just observe when you're in the field. Ask questions and suggest interpretations of the user's actions and motivations. Articulate what matters about the work together.

### 8.1.4 Principle: Good design is systemic

Any good design considers the system and its impact on users as a whole: the handles on a Mini Cooper reflect the aesthetic of the entire car; the iPhone's characteristic user interface elements (including gestures) are carried through the entire design and the apps; all parts of the amazon.com site support the focus on user interests, community ratings, related material, and easy purchase. And all pages of the site look like they are part of the site - a single page could not be changed

Contextual Design provides methods that help a team keep the design coherent. The Contextual Design vision provides a high-level coherent direction; the storyboards provide coherence of task; the User Environment Design ensures structural coherence across the system. All these methods - which are explained in the following section - encourage the designer to think about the entire system, rather than treating each part as its own independent problem to be solved. This provides users with a seamless

Implications for the designer: Use concrete representations to maintain system coherence: function, structure, layout, and flow across the system.

### 8.1.5 Principle: Design depends on explicit representations

When people design, they create physical representations of their concepts. Whether written on the back of a napkin or captured in a high-end modeling tool, designers need a tangible representation of their thoughts. From sketches to formal diagrams, drawings enable designers to work out their ideas, capture their thinking, share it with others, discuss it, and identify weaknesses.

Contextual Design supports this need for a physical representation throughout the design process. Work models make work practice - how users approach their work - explicit, public, and sharable. The User Environment Design shows the structure of the system as experienced by the user. Each technique in Contextual Design has its own tangible representation that supports doing the work, capturing the result, and sharing it with others. These physical representations in Contextual Design are described in the next section.

Implications for the designer: Use drawings, sketches and models to capture key design considerations at every step of the process.

## 8.2 DESCRIPTION OF THE CONTEXTUAL DESIGN PROCESS

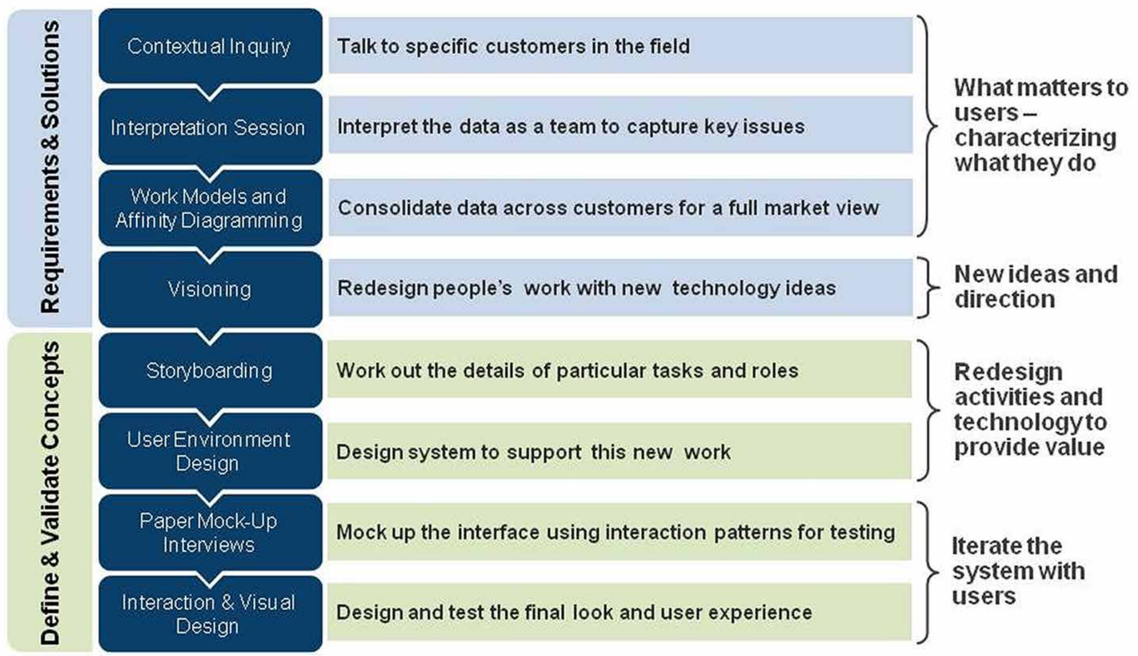



FIGURE 8.1: The Contextual Design Process.



Contextual Design is broadly divided into two major phases (see Figure 8.1). In the following section we'll describe the initial parts of the process, from Contextual Inquiry through visioning. These initial parts are aimed at creating a structured representation of the users' work practice that is actionable for design. Later we'll describe the second phase in the process, which is aimed at working out the details of the design concepts developed in the first half of the process by way of iterative prototyping with users.

### 8.2.1 Contextual Inquiry

The first problem for design is to understand the customers: the people who will use the solution directly (end-users); those who provide them information or use their output (indirect users); those who manage them and are responsible for their success (managers); those who purchase the product and may have their own, quite independent, criteria. For most projects, the main focus is nearly always on the end-users, but it is important to consider and evaluate the needs of the other types of customers as well.

*Contextual inquiry* is an explicit step for understanding who the customers really are and how they work on a day-to-day basis. The difficulty is that, as we described above, work becomes so habitual to end-users that they often have difficulty articulating exactly what they do and why they do it. So the design team conducts one-on-one field interviews with users in their workplace to discover what matters in the work. These are not traditional question and answer interviews. Instead, a contextual interviewer observes users as they work and inquires into the users' actions as they unfold to understand their motivations and strategy. The interviewer and user, through discussion, develop a shared interpretation of the work. It is like an *active inquiry* into the user's world. This inquiry, done in context, is where Contextual Inquiry gets its name.

*Team interpretation sessions* bring a cross-functional design team together to hear the whole story of an interview and capture the insights and learning relevant to their design problem. An interpretation session lets everyone on the team bring their unique perspective to the data, sharing design, marketing, and business implications. Through these discussions, the team comes to understand the customer whose data is being interpreted and their needs, while at the same time capturing issues, drawing work models, and developing a shared understanding of the customer's world.

### 8.2.2 Work Modeling

As described earlier, people's work is complex and full of detail. It's also intangible - there has traditionally been no good way to write down or talk about work practice. Design teams seldom have the critical skill of seeing the structure of work done by others, looking past the surface detail to see the intents, strategies, and motivations that control how work is done - and typical development methodologies do little to encourage this perspective.

Because this is immensely important, so in Contextual Design, *work models* are used to capture the work of individuals and organizations in diagrams. Five different models provide five perspectives on how work is done:

- The *flow model* captures communication and coordination between people to accomplish work. It reveals the formal and informal workgroups and communication patterns critical to doing the work. It shows how work is divided into formal and informal roles and responsibilities.
- The *cultural model* captures culture and policy that constrain how work is done. It shows how people are constrained and how they work around those constraints to make sure the work is done.
- The *sequence model* shows the detailed steps performed to accomplish each task important to the work. It shows the different strategies people use, the intents or goals that their task steps are trying to accomplish, and the problems getting in their way.
- The *physical model* shows the physical environment as it supports or gets in the way of the work. It shows how people organize their environments to make their work easier.
- The *artifact model* shows the artifacts that are created and used in doing the work. Artifacts reveal how people think about their work - the concepts they use and how they organize them to get the work done.

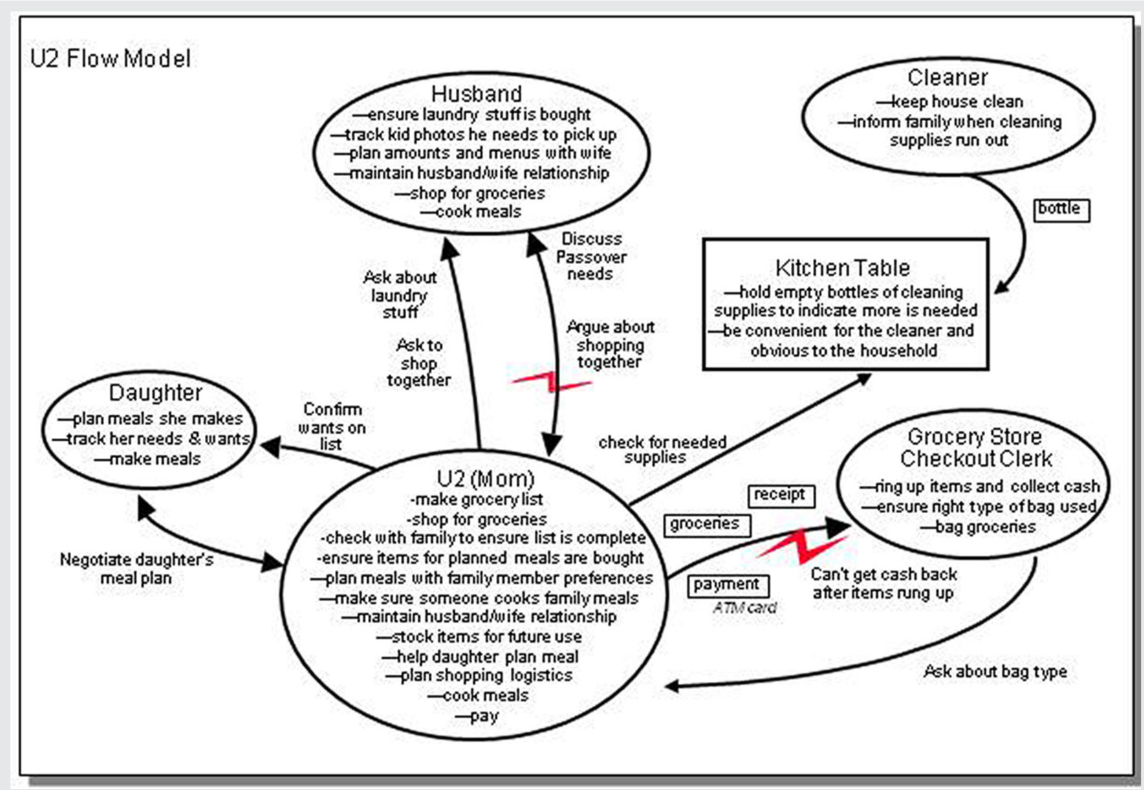



Figure 8.2: **The Flow Model** captures communication and coordination between people to accomplish work. It reveals the formal and informal workgroups and communication patterns critical to doing the work. It shows how work is divided into formal and informal roles and responsibilities.

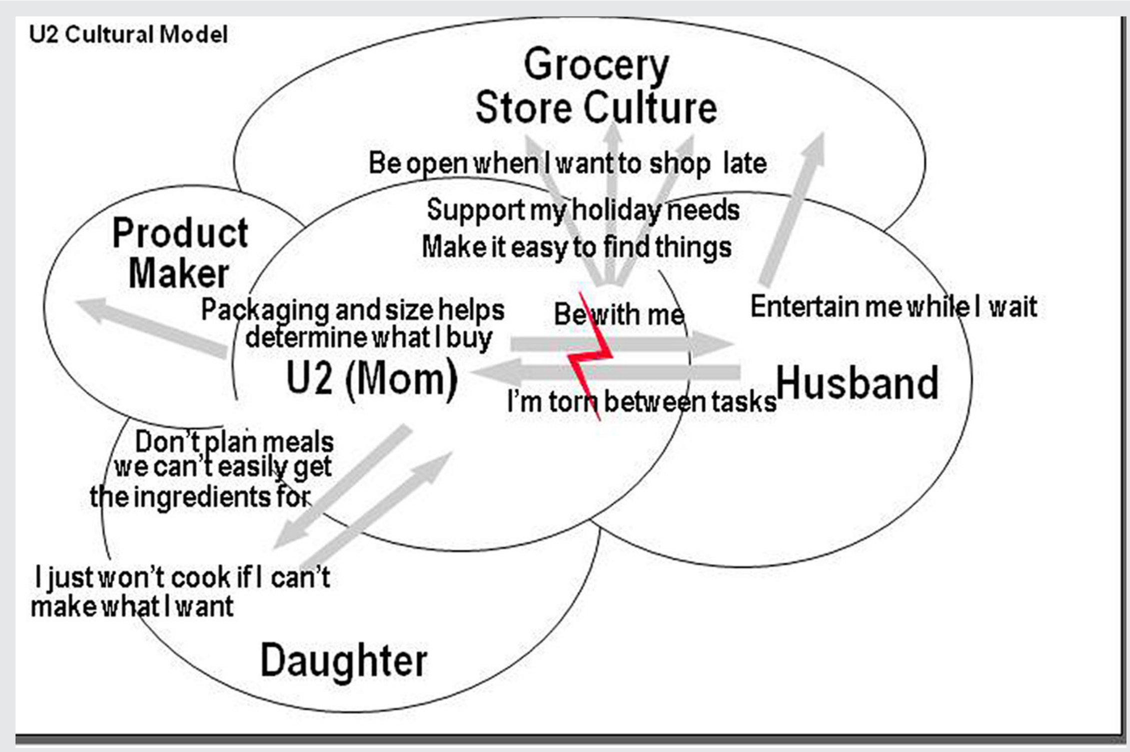



Figure 8.3: **The Cultural Model** captures culture and policy that constrain how work is done. It shows how people are constrained and how they work around those constraints to make sure the work is done.

*Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.*


U2 Sequence -- Grocery Shopping
Pg. 1 of 2
INTENT: Collect groceries needed to feed family and plan what they'll eat
TRIGGER: Weekend and there's time to shop.
Invites husband to go along as a shared activity
INTENT: Balance doing tasks with family time
Husband had 3 other things planned and is overwhelmed
Argument
Husband agrees to come along
Go to favorite grocery store – it's closed
Decide to get only the things that are really needed right now (Passover, breakfast--ricotta, fruit); helper will buy produce later
Go back to second-choice store
Walk to produce
Decide it's no good--will get canned fruit
Decide to go straight to canned fruit instead of walking aisles
H leaves, gets detergent, comes back

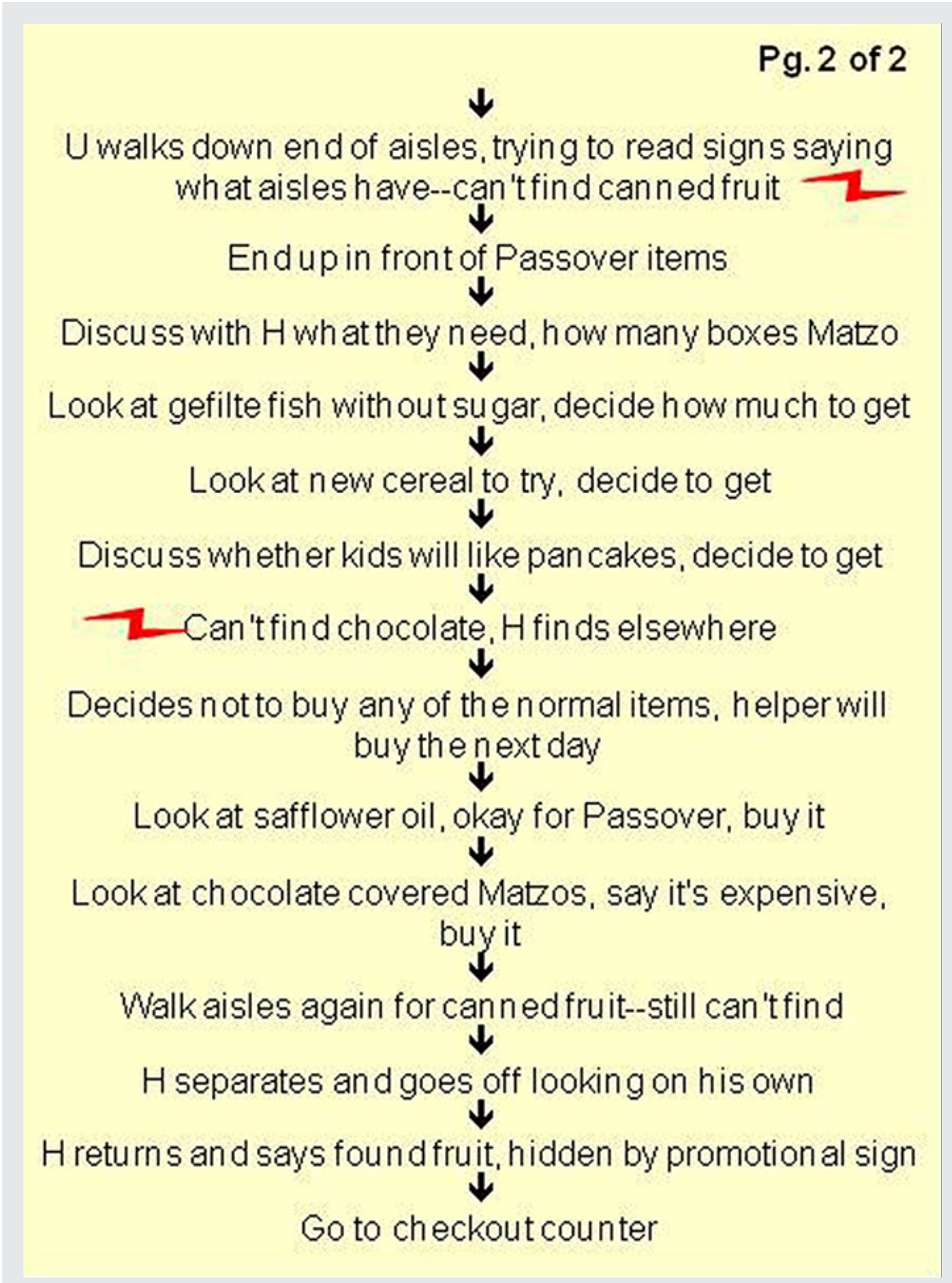



FIGURE 8.4: **The Sequence Model** shows the detailed steps performed to accomplish each task important to the work. It shows the different strategies people use, the intents or goals that their task steps are trying to accomplish, and the problems getting in their way.

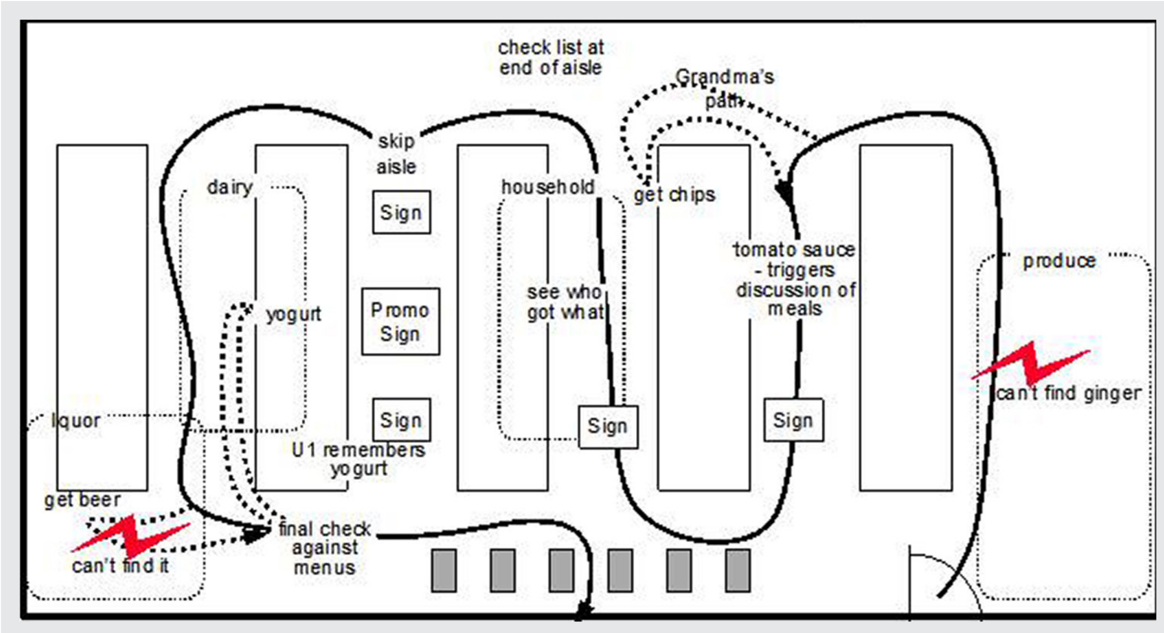



Figure 8.5: **The Physical Model** shows the physical environment as it supports or gets in the way of the work. It shows how people organize their environments to make their work easier.

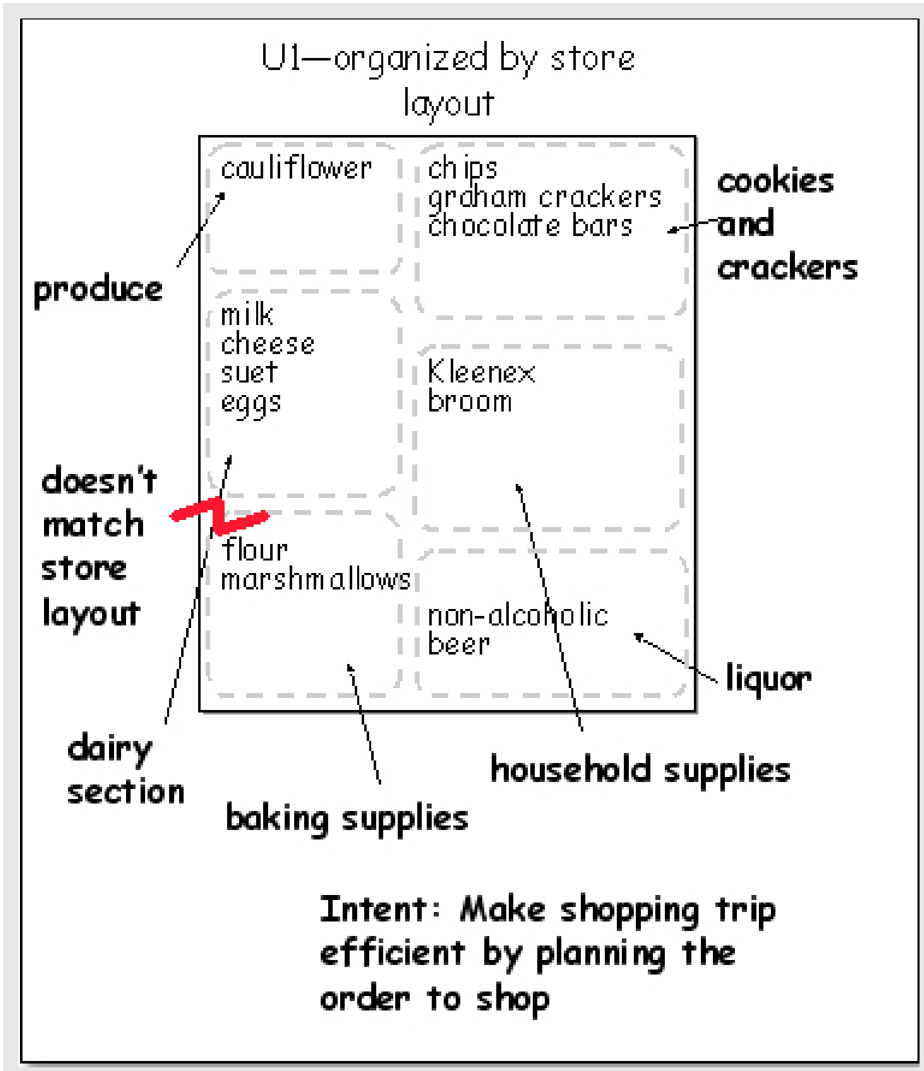



FIGURE 8.6: **The Artifact Model** shows the artifacts that are created and used in doing the work. Artifacts reveal how people think about their work - the concepts they use and how they organize them to get the work done.



### 8.2.3 Consolidation

Systems are seldom designed for a single customer. But designing for a whole *customer population* - the market, department, or organization that will use the system - depends on seeing the common aspects of the work different people do.

*Consolidation* brings data from individual customer interviews together so the team can see common pattern and structure without losing individual variation. The *affinity diagram* brings together issues and insights across all customers into a wall-sized, hierarchical diagram to reveal the scope of the problem.


FIGURE 8.7: **Portion of an Affinity Diagram.** The affinity diagram brings together issues and insights across all customers into a wall-sized, hierarchical diagram to reveal the scope of the problem and the opportunities.

*Consolidated work models* bring together each different type of work model separately, to reveal common strategies and intents while retaining and organizing individual differences. Together, the affinity diagram and consolidated work models produce a single picture of the customer population a design will address. They give the team a focus for the design conversation, showing how the work hangs together rather than breaking it up in lists. They show what matters in the work and guide the structuring of a coherent response, including system focus and features, business actions, and delivery mechanisms.

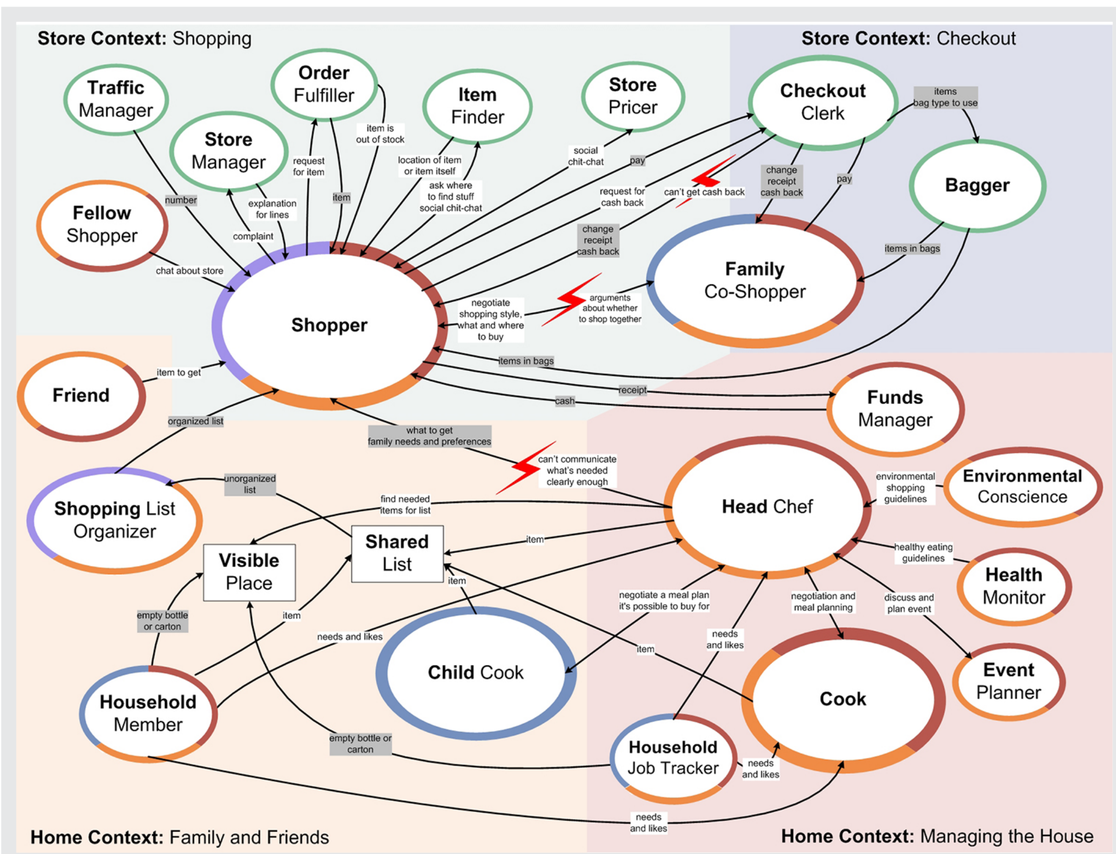



FIGURE 8.8: Consolidated Flow Model.

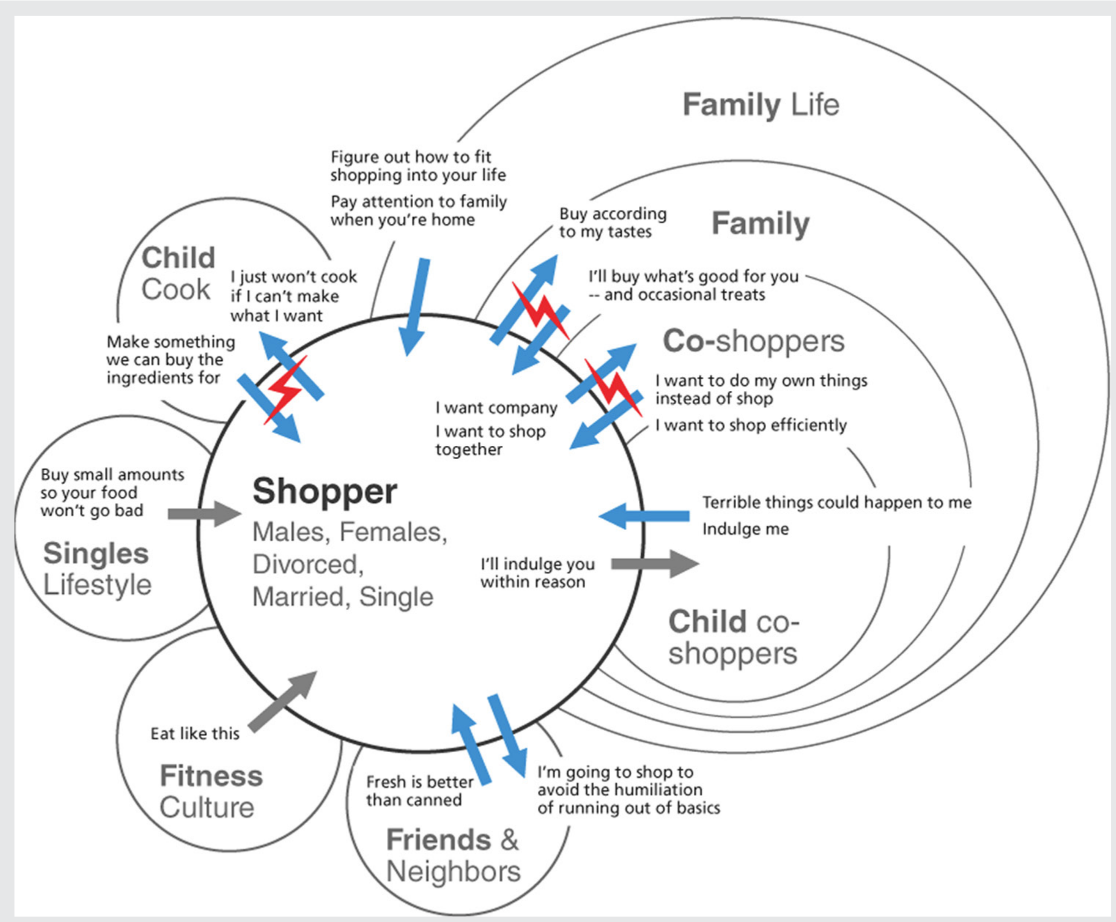



Figure 8.9: Consolidated Cultural Model.

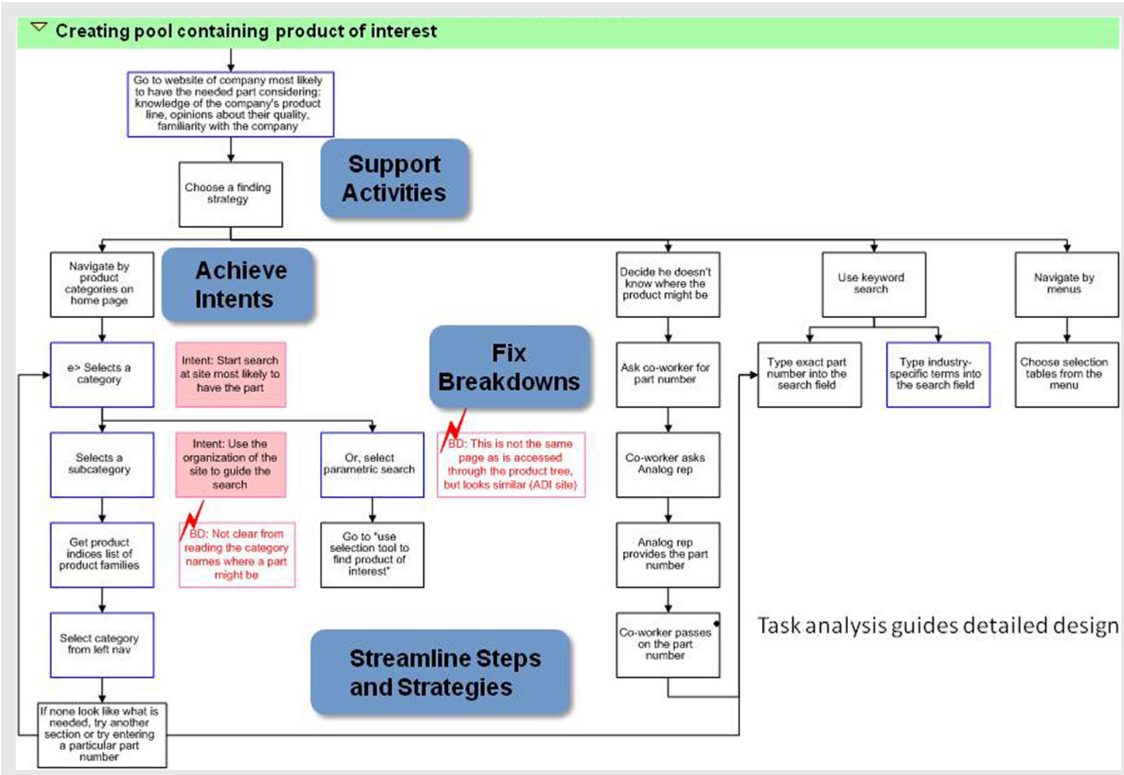



Figure 8.10: Portion of Consolidated Sequence Model.

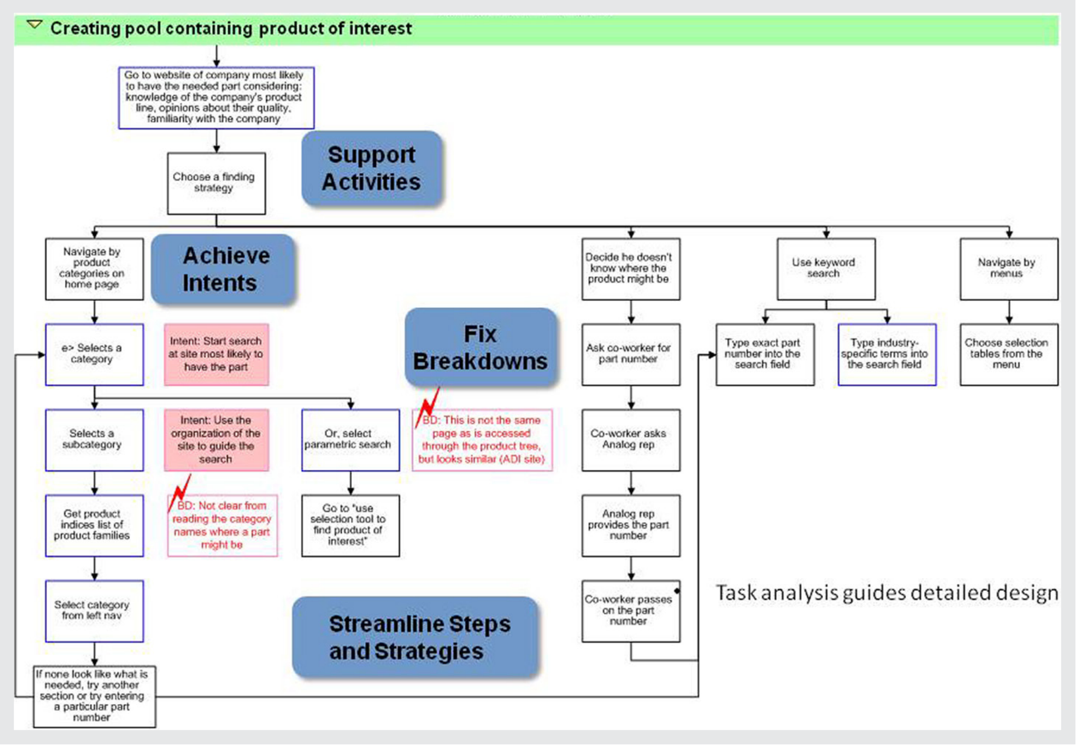
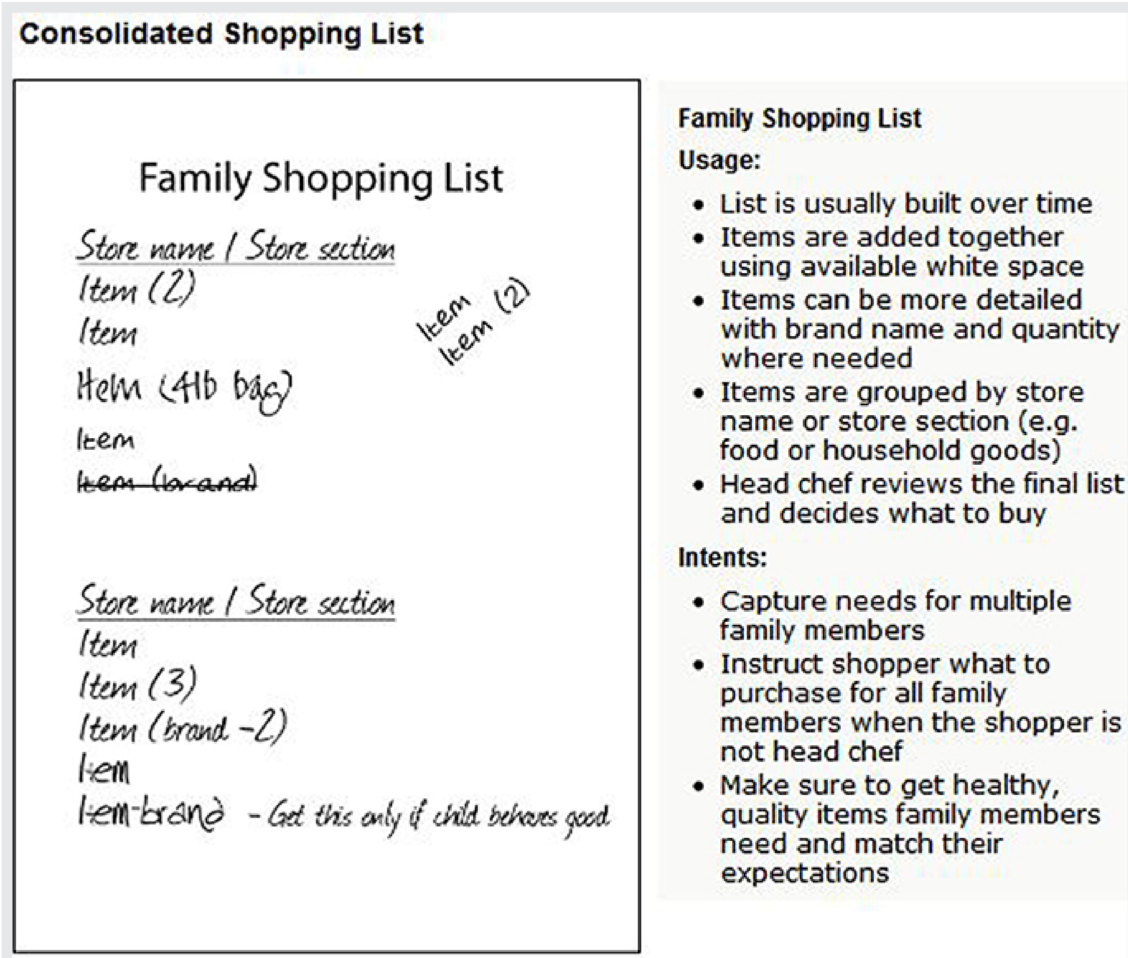



FIGURE 8.11: Consolidated Physical Model.

*Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.*


Consolidated Shopping List
Family Shopping List
Store name / Store section
Item (2)
Item
Item (4lb bag)
Item
Item (brand)
Item
Item (2)
Store name / Store section
Item
Item (3)
Item (brand -2)
Item
Item-brand - Get this only if child behaves good
Family Shopping List
Usage:
• List is usually built over time
• Items are added together using available white space
• Items can be more detailed with brand name and quantity where needed
• Items are grouped by store name or store section (e.g. food or household goods)
• Head chef reviews the final list and decides what to buy
Intents:
• Capture needs for multiple family members
• Instruct shopper what to purchase for all family members when the shopper is not head chef
• Make sure to get healthy, quality items family members need and match their expectations

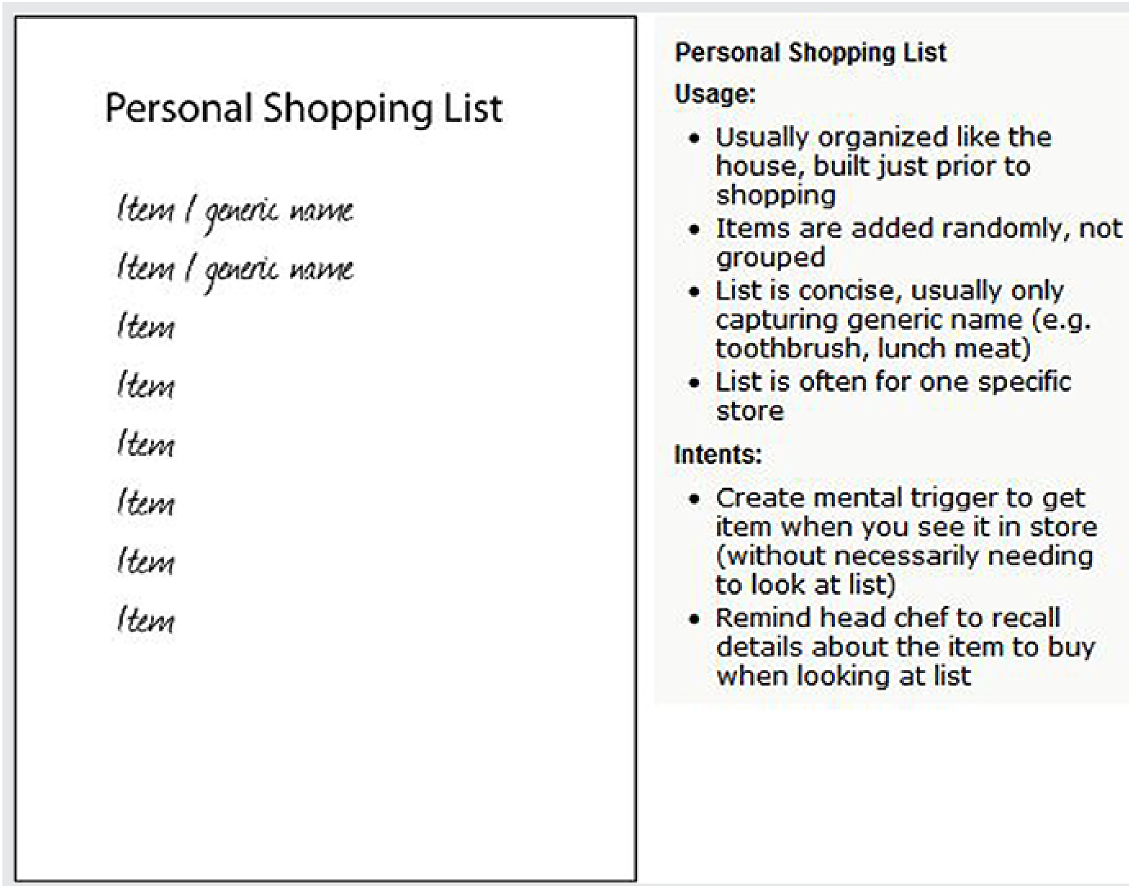



Figure 8.12: Consolidated Artifact Model.

*Courtesy of Sourasith Simonphone. Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.*

### 8.2.4 Personas built with contextual data

Personas can help bring users alive and focus the stakeholders on the relevant issues, if they are built from rich contextual data. Popularized by Alan Cooper, a persona describes typical users of the proposed system as though they were real people (Cooper 1998). Their use is becoming more widespread, though with mixed success. According to Harley Manning's research, "a persona that's not backed by rich contextual data isn't valid, which accounts for much of the mixed success." (Manning 2003)

Contextual Design calls for building personas from the field data the team collected and consolidated to help focus on the characters the design team will vision about in the next step, to help stakeholders segment their market according to practice instead of typical demographics, to clarify branding and prioritization, and to bring the users and their needs to life for developers. Contextual Design personas are built from the detailed data gathered through Contextual Inquiry interviews, so they have the richness and depth needed to drive design.

### 8.2.5 The Design Response: Visioning

Up to this point, a Contextual Design project focuses on understanding the users as they are. Now a team must invent the design solution using technology to transform the tasks, and possibly also designing new business processes to streamline tasks or new services to support the market. A Contextual Design team invents these solutions through *visioning*.

In visioning, the team uses the consolidated data to drive conversations about how to improve users' work by using technology to transform the work practice. This focuses the conversation on how to improve people's lives with technology, rather than on what could be done with technology without considering the impact on peoples' real lives.

The vision captures a story of how customers will do their work in the new world the team invents. A vision includes the system, its delivery, and support structures to make the new work practice successful. It is intentionally rough and high-level - a vision sets a possible design direction, without fleshing out every detail. This enables the team to see the overall structure of the solution and ensure its coherence.

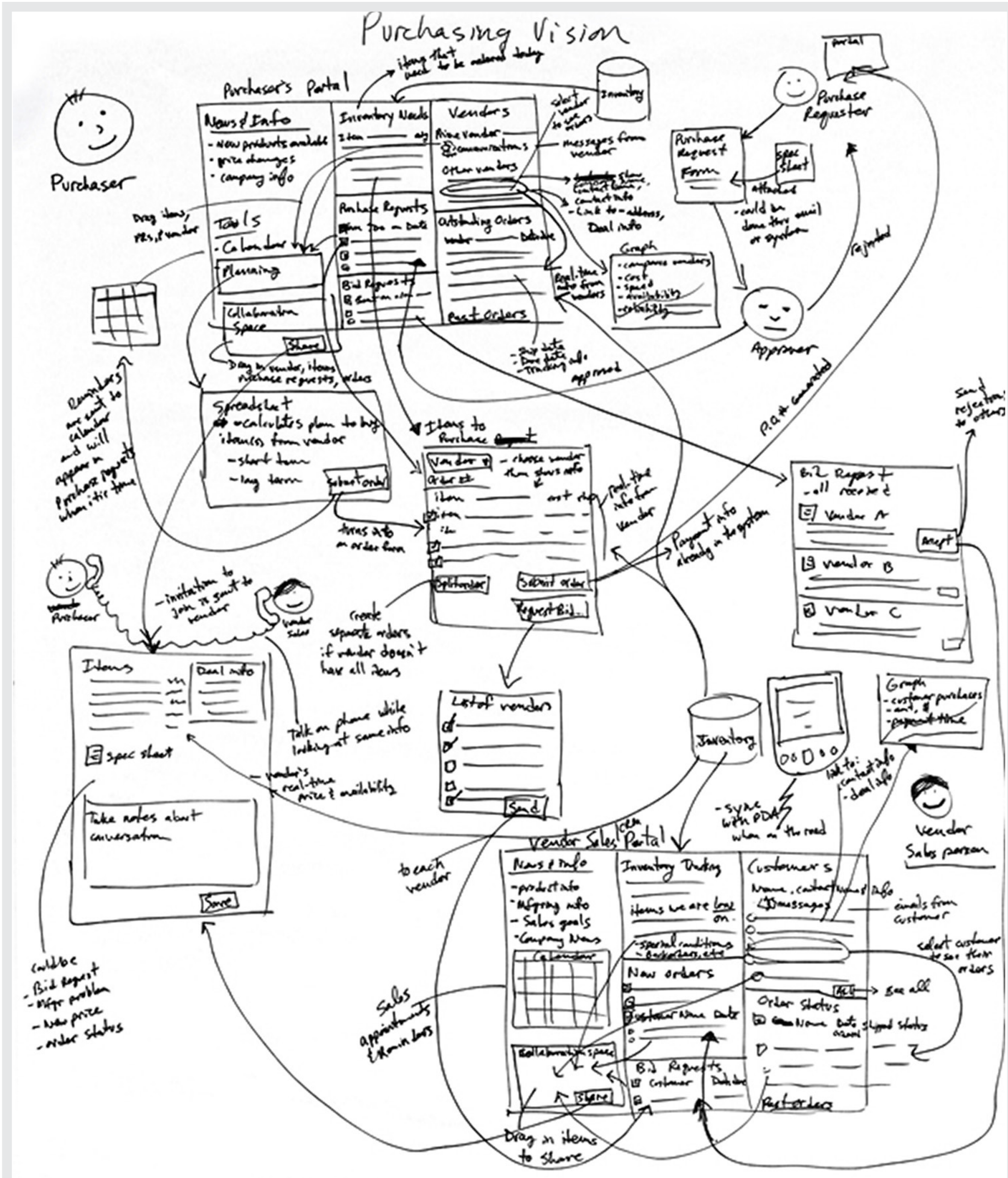



FIGURE 8.13: **The Vision** captures a story of how customers will do their work in the new world the team invents. A vision includes the system, its delivery, and support structures to make the new work practice successful.

### 8.2.6 Storyboards

The vision defines the high-level design response to users' needs. To become actionable, the team must define the detailed function, behavior, and structure of the proposed system. This next level of design must take the users' tasks into account and ensure the right function is defined in the right system places for a smooth workflow. As you'll see in the following section, Contextual Design provides for this structural design through *storyboards* and the *User Environment Design*, and then validates the design through *paper prototypes*.

Each storyboard describes how users will accomplish a task in the new system. They show the steps the user will take and the system function that supports each step. The task may be handed off between users, and may be supported by several systems operating together; the storyboard ensures the task remains coherent across these boundaries.

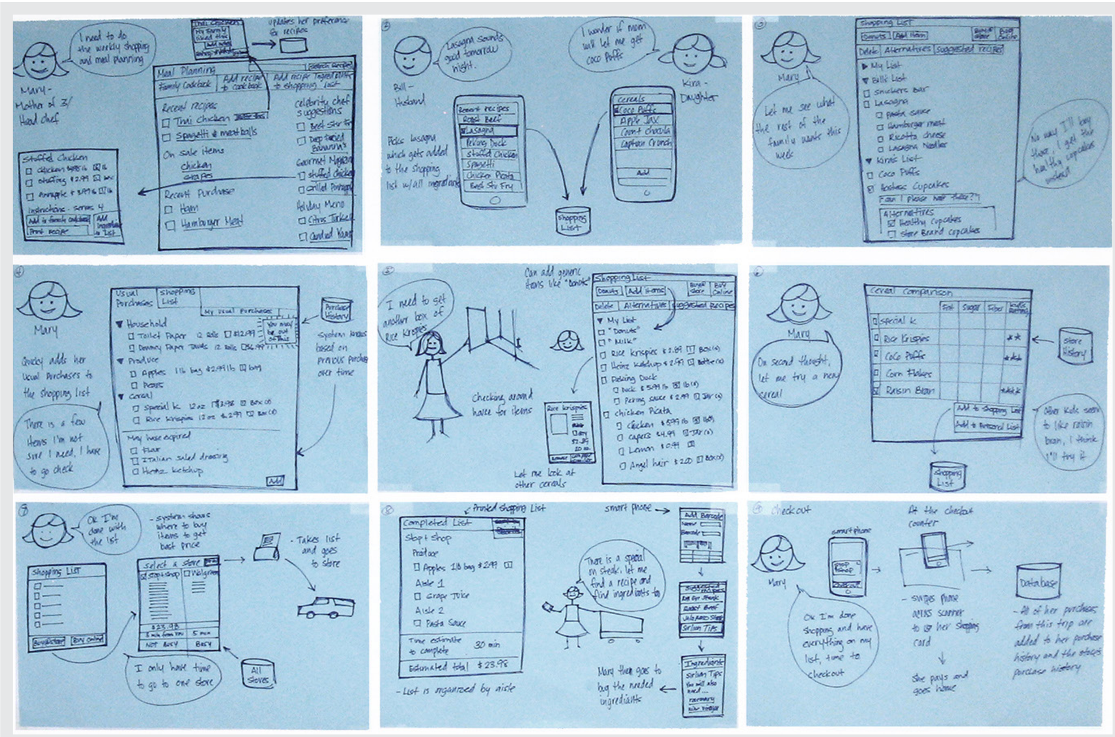



Figure 8.14: **Portion of a Storyboard**. A storyboard is represented as a sequence of "freeze-frame" sketches or cells, each one capturing one step in the overall task.



## 8.2.7 User Environment Design

The storyboards ensure coherence of individual tasks, but the new system must have the appropriate structure to support a natural flow of work through the system no matter what task the user is doing. Just as architects draw floor plans to see the structure and flow of a house, designers need to see the "floor plan" of their new system - the basic structure that will be revealed by the user interface drawing, implemented by an object model, and that responds to the customer work. This "floor plan" is typically not made explicit in the design process.

The *User Environment Design* captures the floor plan of the new system. It shows each part of the system, how it supports the user's work, exactly what function is available in that part, and how the user gets to and from other parts of the system - without tying this structure to any particular user interface.

With an explicit User Environment Design, a team can make sure the structure is right for the user, plan how to roll out new features in a series of releases, and manage the work of the project across engineering teams at a level of abstraction that is above screens and dialogs. Using a diagram which focuses on keeping the system coherent for the user counterbalances other forces that would sacrifice coherence for ease of implementation or delivery.

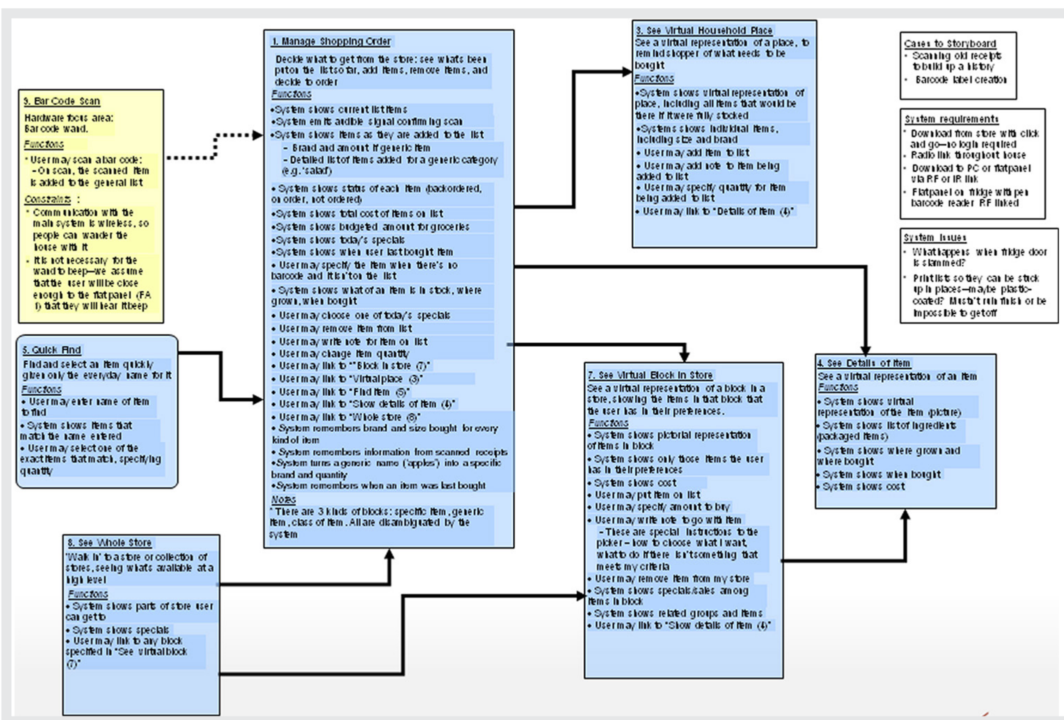



Figure 8.15: **Portion of a User Environment Design**. The User Environment Design shows each part of the system, how it supports the user's work, exactly what function is available in that part, and how the user gets to and from other parts of the system - without tying this structure to any particular user interface.

### 8.2.8 Paper prototyping

Testing is an important part of any system development. It's generally accepted that the sooner problems are found, the less it costs to fix them. So it's important to test and iterate a design early, before anyone gets invested in the design and before spending time writing code. And the simpler a testing process is, the more time is available for multiple iterations to work out the detailed design with users.

*Paper prototyping* develops rough mockups of the system using notes and hand drawn paper to represent windows, dialog boxes, buttons, and menus. The use of paper prototypes is described in many resources, including Carolyn Snyder's book on the subject (Snyder 2003). The design team tests these prototypes with users in their workplace, replaying real work events in the proposed system. When the user discovers problems, they and the designers redesign the prototype together to fit their needs. Rough paper prototypes of the system design test the structure of a User Environment Design and initial user interface ideas before anything is committed to code. Paper prototypes support iteration of the new system, keeping it true to the user needs. Refining the design with users gives designers a customer-centered way to resolve disagreements and work out the next layer of requirements. After several rounds of prototyping, the larger structure of the system design stabilizes. At this point, the design team can continue iterating areas of the user interface.

Once the structure and interaction design are largely stable, the team can develop and test interaction and visual design options with users.

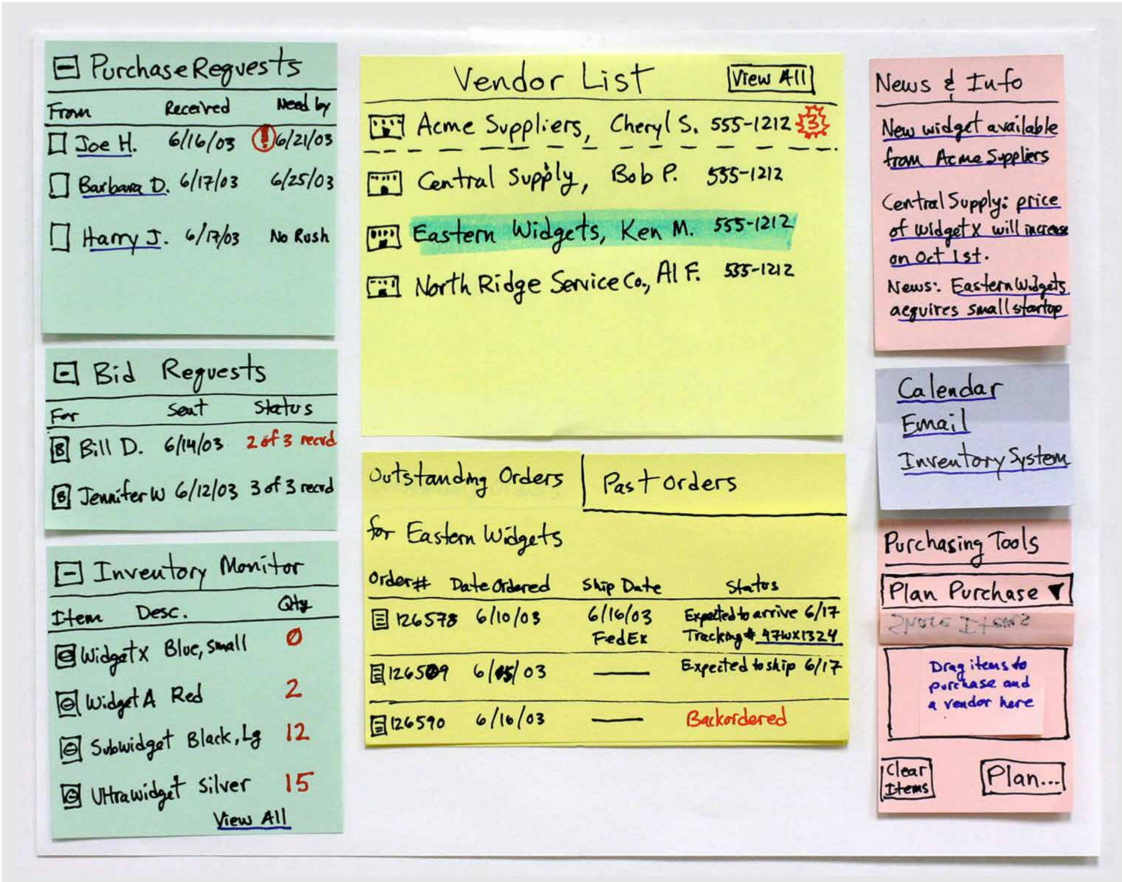



Figure 8.16: **Portion of a Paper Prototype.** Paper prototyping develops rough mockups of the system using notes and hand drawn paper to represent windows, dialog boxes, buttons, menus, and the other user interface elements the customer will use.



## 8.2.9 Driving Product Development

Companies implement a variety of hardware and software development methodologies within which their front-end design process, whether user-centered or not, must fit. Most methodologies define a series of stages, each with deliverables and milestones. Few define specific ways of gathering requirements, instead leaving the specific method open to definition by the product team. Contextual Design is usually included in the requirements gathering step or very early pre-commitment stage gates of these methodologies.

The translation from any kind of research on user needs to design requirements often lacks process and rigor. The structure that Contextual Design offers design teams helps bring some amount of control to this activity. The User Environment Design captures the required function and behavior of the new system, at least for the core work cases. The paper prototypes capture the proposed user interface, though usually only at a rough, wireframe level. These can be harvested to provide the Product Requirements Document and User Interface Specification.

### 8.2.10 Contextual Design and Agile Development

Currently, many organizations are moving to Agile development. In contrast to traditional approaches that emphasize requirements analysis, design, and implementation as distinct phases, Agile methods seek to minimize up-front planning in favor of producing working base levels quickly and often. Feedback from these base levels is used to ensure that the resulting product is useful. Scrum (Schwaber and Beedle 2001) and XP (Beck 2004) (Extreme Programming) are two popular Agile approaches.

Agile development dovetails very nicely with user-centered design (Beyer 2010). But, Agile teams often struggle to include a reliable customer voice, something Agile methods assume they can do. Attempts to substitute stakeholders or internal product owners for the real end-user have only shown how critical that user voice is. Contextual Design provides proven techniques for collecting and using user knowledge which can be adopted by Agile teams.

Before Agile development begins, the initial stages of Contextual Design provide the team with the knowledge they need to write viable user stories. Contextual Inquiry interviews, the affinity diagram, and work models provide the deep understanding of the user needed by the team. Visioning sets the project direction and defines what kind of solution to provide. And storyboards, the User Environment Design, and paper prototypes develop and validate the right function to be included in user stories for Agile release planning. This is critical - paper prototype iterations ensure the team is developing the right design, that it is solving real user problems. It's cheaper and faster to refine the design at this point than in the middle of development iterations.

The key difference between supporting an Agile team and traditional waterfall development is that for an Agile project, the above steps are *all* that need be done. No writing functional specifications, user interface specifications, or architectures. The User Environment Design is kept at the level required for the team to keep its own thinking clear - it is not intended as a communication mechanism to the development team.

Instead, the User Environment Design and paper prototypes are used as the source for writing user stories in the release planning session. They provide enough detail to make it easy to write and estimate stories. Iterations can be planned so each iteration collects stories that, taken together, deliver coherent user value - as defined by the User Environment Design.

During Agile development proper, the techniques of Contextual Design continue to provide critical support to the team. Knowledge gained from field research gives the team confidence in their prioritization of user stories. The detailed user interface can be defined during iterations, usually one iteration ahead of development work. Contextual Inquiry field visits allow detailed user interface designs to be iterated with users. Completed base levels can also be tested using Contextual Inquiry techniques, and the results used to refine the direction of the project.

## 8.3 BACKGROUND AND HISTORY OF CONTEXTUAL DESIGN

Karen Holtzblatt and Hugh Beyer first developed the key parts of the Contextual Design process while working at Digital Equipment Corporation (DEC) in the early 1980s. Karen, a psychologist by training, and Hugh, a developer, recognized the need for a coherent and structured design process that could integrate useful practices from their respective fields, and make it all accessible and actionable to design teams in commercial settings.

Holtzblatt's initial work was a response to the limitations of usability testing and human factors work as it existed in the early 1980's. Whiteside, Bennett, and Holtzblatt (Whiteside et al 1988) introduced and discussed the theoretical foundation for using ethnographic and hermeneutic techniques to understand user practice for the purpose of systems design. At the time, usability methods were focused on lab-based quantitative measures, but these techniques are always limited in the amount of impact they can have and do not lead to wholly new insights and design directions.

Holtzblatt brought techniques from psychology and sociology to the field, showing how the kind of verbal protocol analysis used by Ericsson (Ericsson and Simon 1984) and Piaget (Piaget 1960) could be applied to data collected from users in the field. This data forms the bases for a *grounded theory*, as defined by Glaser and Strauss (Glaser and Strauss 1967), and as such motivates design action. Contextual inquiry was defined as a structured method for gathering and using field data using this theoretical foundation.

The resulting techniques are similar in nature to an ethnographic study. However, contextual inquiry is constrained by the limitations of an engineering project. So field interviews are restricted to a few hours, not days or weeks, and the interaction between interviewer and user is defined as a focused conversation. The purpose of the conversation is to reveal and articulate the nature of the user's work practice, and this purpose is understood and shared by both participants.

At the same time, Holtzblatt was adapting physical mockup techniques developed by Kyng, Ehn and others (Kyng 1988; Ehn 1988) to software. In Denmark, Scandinavian countries mandated that labor representatives be included in any redesign of the workplace by creating mockups of rooms and workstations using large cardboard boxes and other simple, physical representations. Sessions were conducted with the workers in which they ran through typical tasks in his simulated environment, redesigning it to work better as they discovered problems.

Holtzblatt scaled down this method for software, using hand-drawn user interfaces on sticky notes to represent a proposed design and working through the user's own tasks, in their own workplaces, to explore the usefulness of the design. Together, designer and user would modify the prototype in the moment to eliminate problems and add needed function.

Work models were developed by Holtzblatt as a way to capture the discussion in design teams about user work practice - as a way to make elements of work practice explicit to all members of the team. The User Environment Design, similarly, was developed to capture the system structure and function without sidetracking the discussion with user interface details prematurely.

The resulting Contextual Design process was first used at DEC and later gained acceptance in the rapidly growing Human Computer Interaction (HCI) community throughout the 1980's, following the same heretic-to-accepted-practice trajectory that the HCI field itself was undergoing (Caroll 2009). Following a series of articles on various aspects of Contextual Design in the HCI literature (e.g. (Beyer and Holtzblatt 1993), (Holtzblatt and Jones 1993)), the entire process was described in the 1997 text *Contextual Design* (Beyer and Holtzblatt 1997). In 2005, the follow-up handbook *Rapid Contextual Design* (Holtzblatt et al 2005) expanded upon the method and provided more practical guidance. It also addressed an oft-heard criticism that the Contextual Design method could be too labor-intensive or lengthy for some projects.

## 8.4 FUTURE DIRECTIONS

The core elements of Contextual Design have been stable for over a decade and are unlikely to change fundamentally in the future. However, the context in which Contextual Design is used does change and that is likely to drive changes in how the process is used. Here are some possible directions to keep an eye on.

*Agile development*. As Agile processes become more widespread and more accepted, the relationship between Agile development and user-centered processes can be expected to evolve. Agile development itself is strengthened by robust user-centered techniques, but the integration of a coherent design focus with Agile development is still not well-accepted. And the introduction of new Agile methods such as Kanban will continue to provide challenges to good User Experience design.

*Quantitative techniques*. Ideally, the qualitative data provided by contextual inquiry would be augmented with quantitative data provided through research methods such as surveys. When making a business case, it is important to know not just *what* users want, but *how many* potential customers there are and what they are willing to pay for a solution to their problem. Contextual Design can and should be integrated into a whole product concepting and initiation process.

*Enterprise-scale projects*. For large-scale projects, enterprises have to coordinate multiple work streams and hundreds of people over years to accomplish the business goal. Contextual Design can play a key role in identifying the most important problems to solve, prioritizing the rollout of the solution, maintaining coherence of the system vision, and ensuring that as parts are rolled out iteratively the inevitable engineering tradeoffs do not degrade the usability of the system.

## 8.5 WHERE TO LEARN MORE

The definitive sources on Contextual Design are:

**Holtzblatt, Karen and Beyer, Hugh.** *Contextual Design: Defining Customer-Centered Systems*. San Francisco : Morgan Kaufman Publishers, 1997.

**Holtzblatt, Karen, Wendell, Jessamyn and Wood, Shelley.** *Rapid Contextual Design: A How-to Guide to Key Technologies for User-Centered Design*. San Francisco : Morgan Kaufman Publishers, 2005.

**Beyer, Hugh**. *Contextual Design for Agile Teams*. Morgan Claypool. San Rafael, CA. 2010.

Papers and case studies describing uses of Contextual Design abound in the literature. Some have been referenced below, others can be found on the InContext website at: http://www.incontextdesign.com.

## 8.6 COMMENTARY BY JENNIFER J. PREECE

How to cite this commentary in your report

### Jennifer J. Preece

*© Jennifer J. Preece*

Jennifer J. Preece is Dean of the College of Information Studies at the University of Maryland. She researches online communities and is known for her work on what makes such a community successful, and how usability factors interact with sociability in online communities. Preece gained her Ph.D. at the Open University, later becoming faculty there. She went on to be a Research Profess...

Jennifer J. Preece
Jennifer J. Preece is a member of The Interaction Design Foundation

What a wonderful lucid and succinct description of a contextual design. The discussion is focused around a case study with colorful figures to illustrate the step-by-step process that students and those new to the topic will love. Experienced designers too will find material to interest them. For example, there is a discussion about how contextual design practices can be integrated with agile and other methods.

Drs. Karen Holtzblatt and Hugh Beyer also provide a short description of the history of contextual design. It is wisely placed at the end of the article, as many readers will be looking primarily for hands-on advice. But don't overlook this history. It is important for appreciating just how far our discipline has come in integrating users into the design process in a deep and meaningful way that takes account of use contexts, needs, desires and emotions. Karen, a psychologist, and Hugh, a system developer, not only pioneered the development of a new and powerful design methodology, through their work they illustrate the power of interdisciplinary thinking and creativity. Along with co-workers John Whiteside and John Bennett at Digital Equipment Corporation, Karen helped identify the limitations of traditional usability testing (Whiteside et al., 1988). The key one being that while usability testing is good for identifying usability problems that when remedied create incremental improvements, it does not facilitate the large-scale design creativity needed to develop novel systems that offer users an engaging experience.

Contextual Design provided the paradigm shift necessary to create a new kind of design experience, and hence, a new kind of user experience. Gradually over the last twenty plus years contextual design methodology has been refined to provide the rigorous, structured, yet flexible approach described in this article.

Successful methods have two significant characteristics: they are adopted by other researchers and developers, and they can be adapted for use in different situations. Contextual design methodology is widely employed across the world by practitioners and taught to students in human-computer interaction, product design, and related classes (Rogers et al., 2011). I saw an example of the latter first-hand last week while showing a senior administrator around Maryland's ischool. The walls of the hallway were covered with large sheets of paper, marked with colorful markers and adorned with sticky notes – the HCI Masters students were at work! They were engaged in a contextual design exercise under the guidance of Drs. Allison Druin and Karen Holtzblatt. Groups of students were working on different parts of the design, chattering and arguing about where exactly the sticky notes should be placed. The challenge they were set was to develop a system for first-generation college students who may be under-resourced, ethnically diverse, and at times, at-risk.

Allison not only teaches contextual design she has adapted and shaped Karen and Hugh's methodology for her own research on the design of technology for children. Know as "Cooperative Inquiry", Allison brings together teams of adults – researchers, developers, and parents – who work in partnership with children to identify and develop innovative technologies that appeal to children (Druin, 2011). For over fifteen years these intergenerational teams have developed exciting products such as the International Children's Digital Library (www.childrenslibrary.org).

So why has contextual design stood the test of time? There are likely several reasons. First, it was a timely solution to a real problem. Second, it is structured, rigorous and systematic. Third, it respects the needs of real users by enabling them to be partners in the design process. Fourth, it can be adopted and adapted by a wide range of designers from student learners to researchers to professional designers. And fifth, it is challenging and fun!

## References


- Whiteside, J., Bennett, J. and Holtzblatt, Karen (1988): Usability Engineering: Our Experience and Evolution. In: Helander, Martin and Landauer, Thomas K. (eds.). *Handbook of Human Computer Interaction*. North Holland
- Rogers, Y., Sharp, H., Preece, J. (2011) *Interaction Design: Beyond Human Computer Interaction*. 3rd Edition. John Wiley & Sons: Chichester, UK. (See website: id-book.com for interviews with Karen Holtzblatt, which are also contained in the 1st and 2nd editions of the book on pages 313-315 and 578-582 respectively.)
- **Druin**, A. (2011). Children as co-designers of new technologies: Valuing the imagination to transform what is possible. *New directions in youth development: Theory, Practice, and Research: Youth as media creators*, *128*, 35-44.

## 8.7 COMMENTARY BY MARILYN M. TREMAINE

How to cite this commentary in your report

### Marilyn M. Tremaine

*© Marilyn M. Tremaine*

Dr. Marilyn Tremaine is a Research Professor at Rutgers University where she has joint appointments in the College of Communication and Information and the Department of Electrical and Computer Engineering. Prior to Rutgers University, she was a Professor in the Computer Science Department at the University of Toronto and prior to that, a Professor in the University of Michigan Business ...

Marilyn M. Tremaine
Marilyn M. Tremaine is a member of The Interaction Design Foundation

Marilyn is working hard on her commentary. Please check back soon!

## 8.8 COMMENTARY BY DOUGLAS PYLE

How to cite this commentary in your report

### Douglas Pyle

*© Douglas Pyle*

Douglas Pyle is "user experienced". Over the last 12 years he has led UX for web and hardware products, both in the US and Asia, at companies such as Google and Microsoft. He is also an Affiliate Faculty and board member in the Department of Human Centered Design & Engineering at the University of Washington...

Douglas Pyle
Douglas Pyle is a member of The Interaction Design Foundation

Contextual Design is about as close to the customer as you can get.

And for many companies customers are a smelly and scary lot: They talk too much (or not enough), say crazy stuff, and definitely slow things down. The all-too-human side of human factors can be messy, hard, and delay gratification. And by the time we ship the product it's hard to remember how we got here.

With Contextual Design for Agile and stronger UX mindshare across industries, we've gotten over most of these fears now, but they come in new flavors. Today it's scary because if we don't retain control of the innovation process, customers might tell us to build the wrong thing, or worse, build something prosaic/pedestrian.

And we know better than the customer. At least that's what Steve Jobs would say: "It's really hard to design products by focus groups. A lot of times, people don't know what they want until you show it to them."[1] Let's assume that rather than simply the "masses are asses", he means that customers are not good at articulating what they need, which is also a core tenet of Contextual Design. But Jobs implies that first we need something to show users, which would make it mostly conceived and built before customers are involved.

This recent push toward design-led innovation is accompanied by the notion that anything that slows down or pollutes our game-changing design vision is at least extraneous, and at worst severely detrimental to our success in the market. At the core of this debate[2] seems to be the question of locus of innovation: Where will we find this elusive breakthrough? In the customer realm or from the visionary minds within our company? But there is no doubt that the design vision has to come from the company, and any UCD practitioner would tell you that you can't ask customers what the vNext should be. That takes strategy and vision.

IDEO's Tim Brown paints a picture of Design Thinking as a path to product success, and this thinking should gather inspiration from everywhere--including the customer[3]. But Jane Fulton Suri takes it a step further, saying: "Radical innovation requires both evidence and intuition: evidence to become informed, and intuition to inspire us in imagining and creating new and better possibilities."[4]

This is refreshing to hear because although technology has always been transformative, there was a slight naivety to it in the past: products were built to meet a customer need that could usually be articulated, and research methods were very much an exercise in simple requirements and feedback gathering: "What do you want?" and "how are we doing?" Then technology strategy grew to focus on unmet or latent needs, and methods emerged to go a little deeper: site visits to gather requirements and usability studies to see how we were doing.

Now the fish are bigger, and the stakes are higher. The expectation at the outset of new concept development is that the resultant products will actively transform the way people live, and will become their new habit. To be the "architects of the new reality", we need to be thinking much further ahead than where our customers typically focus--in minutiae of their daily lives. As Johan Redstrom would say, we are now trying to *design our users*[5]. But here is the rub: the minutiae of daily human behaviors and life is the only place we will find the seeds of innovation--in those daily experience gaps and latent desires.

Great designers can accomplish much in a design centric company and might even have some big wins. But if the design thinking is not based in deep knowledge of people's lives and context, it will be hard to make products succeed in a repeatable way. Would Amazon attribute the success of the Kindle to their great innovation process, or a great idea with surreptitious market factors?[6]

Newer methods like the design probes used by Philips[7] and Frog, and Richard Zaltman's deep metaphor analysis[8], are attempts to get at these critically competitive morsels: intents, desires, drivers, habits, and practices. Unfortunately many of these methods are not conducted in situ, like Contextual Design.

And when I talk to design researchers at companies like Frog, IDEO, Artefact, or other big thinking consultancies, they are hanging out with the customer. They are living with the customer. They are there not just there to get inspired, or to validate, but to learn something about humans.

People have been studying humans for years, and it takes structure to make sense of the complex interactions and environments in which we live. This is where Contextual Design excels, imbuing the insights with a structure that grounds them, lets them communicate quickly, and helps them live on to inform Big Thing v2.

1. http://www.businessweek.com/1998/21/b3579165.htm
2. http://www.fastcodesign.com/1663220/user-led-innovation-cant-create-breakthroughs-just-ask-apple-and-ikea
3. http://mitworld.mit.edu/video/357/
4. http://www.ideo.com/images/uploads/news/pdfs/Informing_Our_Intuition.pdf
5. http://redstrom.se/johan/abstracts/userdesign.html
6. http://www.freakonomics.com/2009/09/17/survivor-bias-on-the-gridiron/
7. http://www.ijdesign.org/ojs/index.php/IJDesign/article/view/305/242
8. http://en.wikipedia.org/wiki/Zaltman_Metaphor_Elicitation_Technique; or http://www.olsonzaltman.com/html/howzmet.html

## 8.9 REFERENCES

*Proceedings of the 33rd SIGCSE technical symposium on Computer science education 2002.*

*JCDL05 Proceedings of the 5th ACM/IEEE-CS Joint Conference on Digital Libraries 2005.*

*Armstrong, Anne-Marie (2004): Instructional Design in the Real World: A View from the Trenches. Idea Group Publishers*

*Beck, Kent (2004): Extreme Programming Explained: Embrace Change (2nd Edition). Addison-Wesley Professional*

*Beck, Kent (1999): Extreme Programming Explained: Embrace Change. Reading, MA, Addison-Wesley Publishing*

*Beyer, Hugh R. (2010a): Contextual Design for Agile Teams. Morgan Claypool*

Beyer, Hugh R. (2010b): User-Centered Agile Methods. Morgan and Claypool Publishers

Beyer, Hugh R. and Holtzblatt, Karen (1998): Contextual Design: Defining Customer-Centered Systems. San Francisco, Morgan Kaufmann Publishers

Beyer, Hugh R. and Holtzblatt, Karen (1997): Contextual Design: Defining Customer-Centered Systems (Interactive Technologies). Morgan Kaufmann

Beyer, Hugh R. and Holtzblatt, Karen (1993): Contextual Design: Toward a Customer-Centered Development Process. In: Software Development 93 Spring Proceedings February, 1993, Santa Clara, CA, USA.

Beyer, Hugh R. and Holtzblatt, Karen (1995): Apprenticing with the Customer. In Communications of the ACM, 38 (5) pp. 45-52

Carroll, John M. (2012). Human Computer Interaction - brief intro. Retrieved 6 June 2013 from [URL to be defined - in press]

Cooper, Alan (1998): The Inmates Are Running the Asylum: Why High Tech Products Drive Us Crazy and How to Restore the Sanity. Sams Publishing

Cooper, Alan (1999): The Inmates Are Running the Asylum: Why High Tech Products Drive Us Crazy and How to Restore the Sanity. Sams

Curtis, Paula, Heiserman, Tammy, Jobusch, David, Notess, Mark and Webb, Jayson (1999): Customer-Focused Design Data in a Large, Multi-Site Organization. In: Altom, Mark W. and Williams, Marian G. (eds.) Proceedings of the ACM CHI 99 Human Factors in Computing Systems Conference May 15-20, 1999, Pittsburgh, Pennsylvania. pp. 608-615

Ehn, Pelle (1988): Work-Oriented Design of Computer Artifacts. Stockholm, Arbetslivscentrum

Ericsson, K. A. and Simon, Herbert A. (1984): Protocol Analysis: Verbal Reports as Data. Cambridge, MA, MIT Press

Garfinkel, Harold (1967): Studies in Ethnomethodology. London, United Kingdom, Polity Press

Glaser, Barney and Strauss, Anselm (1967): The Discovery of Grounded Theory: Strategies for Qualitative Research. Aldine Transaction

Holtzblatt, Karen and Beyer, Hugh R. (1993): Making Customer-Centered Design Work for Teams. In Communications of the ACM, 36 (10) pp. 92-103

*Holtzblatt, Karen and Jones, Sandra (1993): Contextual Inquiry: A Participatory Technique for System Design. In: Namioka, Aki and Schuler, Doug (eds.). "Participatory Design: Principles and Practice". Hillsdale, NJ, USA: Lawrence Earlbaumpp. 177-210*

*Holtzblatt, Karen, Wendell, Jessamyn Burns and Wood, Shelley (2004): Rapid Contextual Design: A How-to Guide to Key Techniques for User-Centered Design (Interactive Technologies). Morgan Kaufmann*

*Kyng, Morten (1988): Designing for a Dollar a Day. In: Greif, Irene (ed.) Proceedings of the 1988 ACM conference on Computer-supported cooperative work September 26 - 28, 1988, Portland, Oregon, United States. pp. 178-188*

*Lárusdóttir, M. K. "Using Rapid Contextual Design at Reykjavik University", accepted paper for the workshop named "HCIEd.2006-1 inventivity: Teaching theory, design and innovation in HCI" held by British Computer Society HCI Group (BHCIG), the International Federation of Information Processing (IFIP) WG13.1 Education, and the Irish Computer Society" in Limrick in Ireland, 23rd – 24th March, 2006*

*Manning, Harley (2003). The Power Of Design Personas. Retrieved [Date unavailable] from Forrester Research: http://www.forrester.com/ER/Research/Report/Summary/0,1338,33033,00.html*

*Mcdonald, Sharon, Monahan, Kelly and Cockton, Gilbert (2006): Modified contextual design as a field evaluation method. In: Proceedings of the Fourth Nordic Conference on Human-Computer Interaction 2006. pp. 437-440*

*Notess, Mark (2005): Using Contextual Design for Digital Library Field Studies. In: JCDL05 Proceedings of the 5th ACM/IEEE-CS Joint Conference on Digital Libraries 2005.*

*Notess, Mark (2004): Applying Contextual Design to Educational Software Development. In: Armstrong, Anne-Marie (ed.). "Instructional Design in the Real World: A View from the Trenches (Advanced Topics in Information Resources Management)". Idea Group Publishers*

*Piaget, Jean (1960): The Child's Conception of the World. Rowman and Littlefield Publishers, Inc*

*Polanyi, Michael (1958): Personal Knowledge: towards a post-critical philosophy. London, Routledge*

*Rockwell, Chris (1999): Customer connection creates a winning product: building success with contextual techniques. In Interactions, 6 (1) pp. 50-57*

*Schuler, Douglas and Namioka, Aki (eds.) (1993): Participatory Design: Principles and*

*Practices. Hillsdale, NJ, USA, Lawrence Erlbaum Associates*

Schwaber, Ken and Beedle, Mike (2001): *Agile Software Development with Scrum. Prentice Hall*

Snyder, Carolyn (2003): *Paper Prototyping: The Fast and Easy Way to Design and Refine User Interfaces. Morgan Kaufmann*

Weinberg, Jerry B. and Stephen, Mary L. (2002): *Participatory design in a human-computer interaction course: teaching ethnography methods to computer scientists. In: Proceedings of the 33rd SIGCSE technical symposium on Computer science education 2002. pp. 237-241*

Whiteside, John, Bennett, John and Holtzblatt, Karen (1988): *Usability Engineering: Our Experience and Evolution. In: Helander, Martin and Landauer, Thomas K. (eds.). "Handbook of Human Computer Interaction". North Holland*

## About the authors

### Karen Holtzblatt

*© Karen Holtzblatt*

Has also published under the name of:

"Karen A. Holtzblatt"

PERSONAL HOMEPAGE:

incontextdesign.com/people/karen-holtzblatt/

CURRENT PLACE OF EMPLOYMENT:

InContext

Recognized as a leader in the design community, Karen has pioneered transformative ideas and design approaches throughout her career. Karen is the inventor of Contextual Inquiry—the industry standard for gathering field data to understand how technology impacts the way people work. Contextual Inquiry and the design processes based on it provide a revolutionary approach for designing new products and systems based on a deep understanding of the context of use. Contextual Inquiry forms the base of Contextual Design, InContext's full customer-centered design process. Karen co-founded InContext Enterprises in 1992 to use Contextual Design techniques to coach product teams and deliver customer-centered designs to businesses across multiple industries. The books, Contextual Design: Defining Customer Centered Systems, and Rapid Contextual Design, are used by companies and universities all over the world. Karen is a member of the CHI Academy (awarded to significant contributors in the Computer Human Interaction Association) and received the first Life Time Award for Practice at CHI2010 for her contributions to the field. Karen's extensive experience with teams and all types of work and life practice underlies the innovation and reliable quality consistently delivered by InContext's teams. Karen also has more than 20 years of teaching experience, professionally and in university settings. She holds a doctorate in applied psychology from the University of Toronto.

## Hugh R. Beyer

© *Hugh R. Beyer*

Has also published under the name of:

"Hugh Beyer"

Personal Homepage:

incontextdesign.com/people/hugh-beyer/

Current place of employment:

InContext

Hugh Founder and CTO of InContext. He has more than 20 years of experience building and designing applications, systems, and tools. Before co-founding InContext, Hugh acted as lead developer and architect in a range of systems at Digital Equipment Corp. His domains of experience include object-oriented repositories, databases, and integrated software development environments. Since starting InContext, Hugh has overseen the design of applications from desktop to web to mobile, and from enterprise to small business to consumers in the wide variety of industries supported by InContext. He holds a B.S. degree in Applied Mathematics from Harvard.

## Your notes and thoughts on chapter 8

Record your notes and thoughts on this chapter. If you want to share these thoughts with others online, go to the bottom of the page at:
http://www.interaction-design.org/encyclopedia/contextual_design.html

**NOTES**

# Chapter 9

# Mobile Computing

*by Jesper Kjeldskov.*

One of the things that makes mobile computing an interesting topic of research and design is that the area is strongly driven by innovation, characterised by rapidly evolving use, and has enormous market potential and growth. New technologies are constantly being developed, new use domains are constantly being explored, and successful new ideas and applications reach millions of users. In fact, by the end of 2010 more smartphones than personal computers were, for the first time, being sold worldwide, with more than 100 million units shipped in the last three months of that year alone. Reflecting this dynamic and rapidly evolving nature of the area, the industrial lead position has been passed on several times within only a decade, from Palm to Nokia to Apple, and is most likely to be passed on again in the future. This obviously motivates researchers and designers to keep innovating and developing new technology and applications. A primary driver of mobile technology development has been the enormous uptake of interactive systems and devices for work as well as for leisure. Mobile phones have long been something almost everyone owns at least one of and uses extensively for personal purposes and not just for work. With Internet and multimedia-enabled phones such as the Apple iPhone, smart phones have now firmly reached this mass market too and are no longer something exclusively for a small elite of business professionals. The uptake of mobile technology in our work and private spheres has had a huge impact on the way we perceive and use these technologies. They are no longer just computers on batteries. They have become functional design objects, the look, feel and experience of which we care deeply about, and that we juggle in multitude in our everyday lives.

## 9.1 INTRODUCTION

Mobile computing is a relatively new field of research with little more than three decades of history. During its lifetime, it has expanded from being primarily technical to now also being about usability, usefulness, and user experience. This has led to the birth of the vibrant area of mobile interaction design at the intersections between, among others, mobile computing, social sciences, human-computer interaction, industrial design, and user experience design. Mobile computing is a significant contributor to the pervasiveness of computing resources in modern western civilisation. In concert with the proliferation of stationary and embedded computer technology throughout society, mobile devices such as cell phones and other handheld or wearable computing technologies have created a state of ubiquitous and pervasive computing where we are surrounded by more computational devices than people (Weiser 1991). Enabling us to orchestrate these devices to fit and serve our personal and working lives is a huge challenge for technology developers, and "as a consequence of pervasive computing, *interaction design* is poised to become one of the main liberal arts of the twenty-first century" (McCullough 2004).

The field of mobile computing has its origin in a fortunate alignment of interests by technologists and consumers. Since the dawn of the computing age, there have always been technological aspirations to make computing hardware smaller, and ever since computers became widely accessible, there has been a huge interest from consumers in being able to bring them with you (Atkinson 2005). As a result, the history of mobile computing is paved with countless commercially available devices. Most of them had short lifespan and minimal impact, but others significantly pushed the boundaries of engineering and interaction design. It is these devices, and their importance, that I wish to emphasize here.

## 9.2 SEVEN WAVES OF MOBILE COMPUTING

The history of mobile computing can be divided into a number of eras, or waves, each characterised by a particular technological focus, interaction design trends, and by leading to fundamental changes in the design and use of mobile devices. In my view, the history of mobile computing has, so far, entailed seven particularly important waves. Although not strictly sequential, they provide a good overview of the legacy on which current mobile computing research and design is built.

1. Portability
2. Miniaturization
3. Connectivity
4. Convergence
5. Divergence
6. Apps
7. Digital ecosystems

The era of focus on *Portability* was about reducing the size of hardware to enable the creation of computers that could be physically moved around relatively easily. *Miniaturization* was about creating new and significantly smaller mobile form factors that allowed the use of personal mobile devices while on the move. *Connectivity* was about developing devices and applications that allowed users to be online and communicate via wireless data networks while on the move. *Convergence* was about integrating emerging types of digital mobile devices, such as Personal Digital Assistants (PDAs), mobile phones, music players, cameras, games, etc., into hybrid devices. *Divergence* took an opposite approach to interaction design by promoting information appliances with specialised functionality rather than generalized ones. The latest wave of *apps* is about developing matter and substance for use and consumption on mobile devices, and making access to this fun or functional interactive application content easy and enjoyable. Finally, the emerging wave of *digital ecosystems* is about the larger wholes of pervasive and interrelated technologies that interactive mobile systems are increasingly becoming a part of.

### 9.2.1 Portability

The first mobile computers, the precursors to present time's laptops, were developed in the late 1970s and early 1980s inspired by the portability of Alan Kay's Dynabook concept from 1968 (Kay 1972). The Dynabook concept was originally thought of as a machine for children, but observant entrepreneurs, such as the founder of GRiD Systems, John Ellenby, quickly realised that the starting point for something that innovative would have to be "the customer with the most money and the most demanding need" (Moggridge 2007).

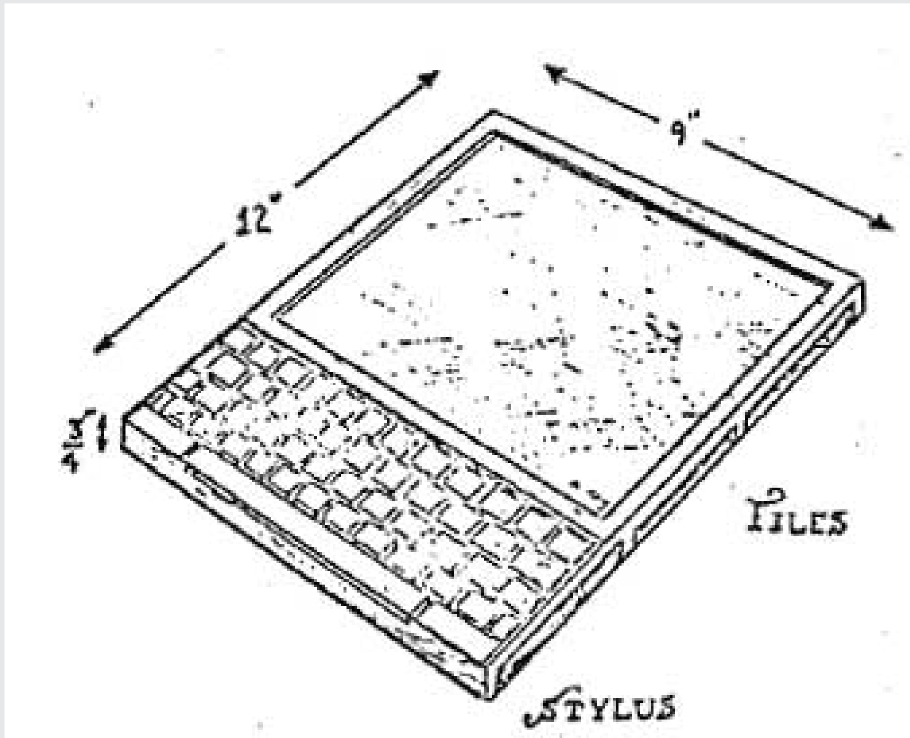



Figure 9.1: Alan Kay's Dynabook: 'a personal computer for children of all ages' (Kay 1972).



The first laptop computer was the GRiD Compass 1101 designed by Bill Moggridge as early as 1981 in response to the design brief of fitting within half the space of a briefcase (Moggridge 2007; Atkinson 2005). The Compass had a 16MHz Intel 8086 processor, 256K DRAM, a 6-inch 320x240 pixel flat screen display, 340kb bubble memory, a 1200 bit/s modem, weighed 5 kg, and ran its own graphical operating system called GRiD OS. It was primarily sold to the U.S. government and was, amongst others, used by NASA on Space Shuttle missions during the early 1980s, and in combat. The GRiD Compass featured a stunning forty-three innovative features in its utility patent, including the flat display and hinged screen. The first portable computer to reach real commercial success, however, was the suitcase-style Compaq Portable from 1982, which as the first official IBM clone could run MS-DOS and standard PC programs. In 1988, Grid Systems also developed the first tablet computer, the GRiDpad, initiated and led by Jeff Hawkins who later designed the first PalmPilot and founded Palm Computing.

In terms of design longevity and impact, Bill Moggridge's work on the first laptop computer and Jeff Hawkins' work on the GRiDpad illustrates the value of careful and well-considered interaction design in mobile computing. The GRiD Compass was superior in terms of its design and performance for a decade. It defined the folding design still used in today's laptops 30 years later, and its basic form factor was not surpassed until the Apple PowerBook 100 introduced the, now standard, clam-shell design and integrated pointing device in 1991. The basic design of the GRiDpad paved the way for tablet computers and handheld devices such as the Apple Newton, the PalmPilot, and even the iPad.


*Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.*

*Copyright © Compaq Computers. All Rights Reserved. Used without permission under the Fair Use Doctrine (as permission could not be obtained). See the "Exceptions" section (and subsection "allRightsReserved-UsedWithoutPermission") on the page copyright notice.*




Figure 9.2 A-B-C: Mobile computers in the 1980-90s: GRiD Compass 1101 (1981), Compaq Portable 1 (1982), and GRiDpad 1910 (1989).

### 9.2.2 Miniaturization

By the early 1990s, the size of computer hardware had reached a point that allowed radically new and smaller form factors of mobile computers to evolve and emerge on the market. These predominantly handheld devices were labelled palmtop computers, digital organizers, or “Personal Digital Assistants” (PDAs). PDAs differed from laptop PCs by being truly mobile and something that the users could operate while actually moving around physically. They were not thought of as alternatives to desktop or laptop computers, but rather as small and lightweight supplemental devices for busy businessmen spending some of their time away from their PC. The first PDA was the Apple Newton from 1992. In 1997, the first PalmPilot was introduced, and in 2000 Compaq released the iPAQ Pocket PC. Whereas the focus of laptop computing was predominantly on portability and mobile access to documents and applications available on desktop computers, palmtop computing introduced an additional focus on applications and interaction styles designed specifically for mobile devices and mobile users.

The PDA generation of mobile devices represented a number of distinct interaction design choices and form factors. Most notably, they introduced the combination of a relatively small touch-sensitive screen and a separate pen (or stylus) for user interaction. Using the stylus, the user could interact with content directly on the screen and enter text via an on-screen keyboard or through handwriting recognition software. Other interaction design innovations included function buttons for accessing pre-defined applications and functions, navigation keys for operating menus, and the “one-click” dock for synchronizing with a stationary computer and for charging. While the Psion series 3 to 5 replicated a “laptop in miniature” design, the Newton, PalmPilot and iPAQ all represented a fundamentally new mobile computing form factor where the majority of the device’s surface was used for its display. In terms of interaction design, the PalmPilot in particular was a product of careful and detailed rethinking of the emerging class of handheld computers; what they should look and feel like, what functions they should perform, and how they should perform them. As an example, the creator of the PalmPilot, Jeff Hawkins, later explained how he carried blocks of wood with him in different sizes and shapes until he had reached the perfect physical form for the device (Bergman and Haitani 2000).


Dori,
The design looks good.
Could you please add our
logo to the top left corner.
Thanks,
Kevin


Palm Pilot
Battery
pilot


Angebote
Briefe
Privates
Budget
Daten
Termine
Meine Unterschrift
Sprachnotiz
Notiz
S5MXpro.htm
SHELL
Word
Sheet
Contacts
Agenda
Email
Calc
Jotter
Extras




Figure 9.3 A-B-C-D: Mobile computers in the 1990-00s: Apple Newton (1992), PalmPilot (1997), Psion 5 (1997), and Compaq iPAQ (2000).

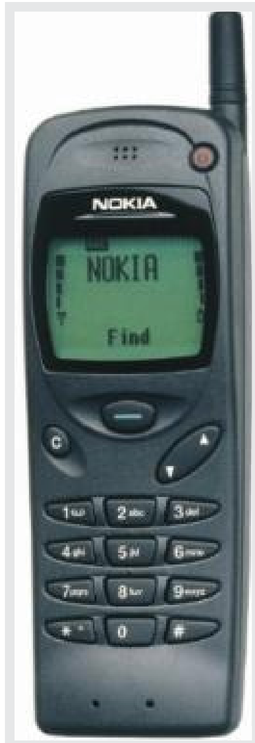
With the emergence of PDAs came also new categories of applications developed specifically for mobile devices and users. The devices each had their own operating systems, optimized for their particular screen sizes and input capabilities, and a suite of standard applications for calendars, contacts, note taking, and email. Adding to this, a wide range of 3rd party applications soon became available for purchase or, as something new, downloadable via the Internet. By the late 1990s, application development specifically for mobile devices was an acknowledged research area and profession, and in 1998 the first international workshop on Human-Computer Interaction with Mobile Devices (Mobile HCI'98) was held in Glasgow specifically addressing the emerging challenge of interaction design and user experiences for mobile devices, systems and services.

### 9.2.3 Connectivity

The third wave of mobile computing had its origins in wireless telecommunication. As early as 1973, a Motorola team led by Martin Cooper developed and patented a handheld mobile phone concept that led to the first commercial mobile phone small enough to be carried, the DynaTAC 8000X, in 1983.

FIGURE 9.4: The first handheld cell phone: Motorola DynaTAC 8000X (1983).

In the 1980s and early 1990s, mobile phones were not really considered to be computers. However, with the introduction of the digital Global System for Mobile Communications (GSM) mobile phone system in 1991, which also included the Short Message Service (SMS) communication component, the complexity and functionality of handsets began evolving rapidly. So did the uptake of mobile phone technology by the broad population worldwide. This meant that mobile phone developers were suddenly faced with a huge challenge of interaction design not only for making phone calls, but also for handling contacts, calendars, text-based messages, and browsing the Internet. In the late 1990s, interaction design for mobile phones was unarguably dominated by the work at Nokia, which led to a series of groundbreaking handsets. The challenges of the time were to design for tiny low-resolution displays and for input capabilities limited to a 12-key numeric keypad alongside a small number of function and navigation keys. One of the first mobile phones explicitly resulting from a careful process of interaction design in the 1990s was the Nokia 3110. It introduced a simple graphical menu system and the “Navi-key” concept for simplifying user interaction — an interaction design that reached the hands of more than 300 million users through subsequent Nokia handsets (Lindholm and Keinonen 2003). In 1999, the basic interaction design of the Nokia 3110 was extended with T9 predictive text for SMS messaging, games, customisable ring tones, and changeable covers for the extremely successful Nokia 3210.


NOKIA
NOKIA
Find

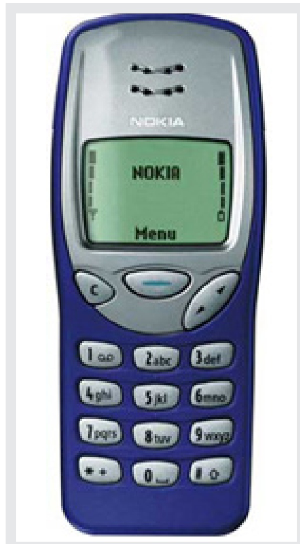






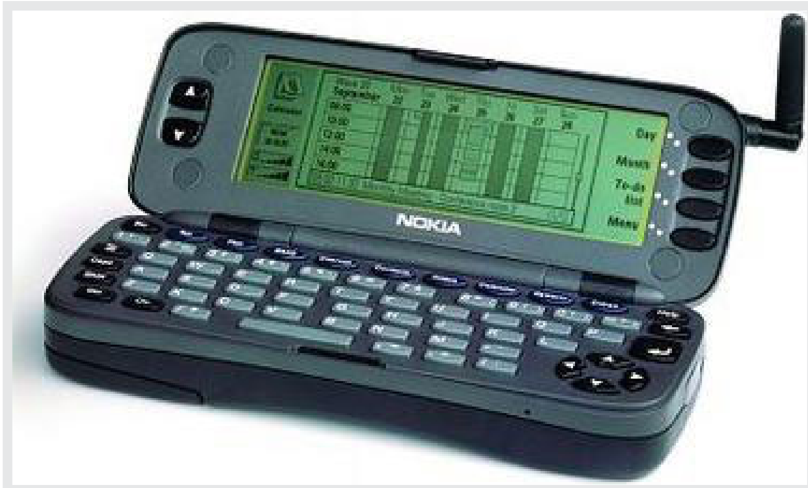


Figure 9.5 A-B-C: Three mobile interaction design milestones: Navi-key, T9, and WAP: Nokia 3110 (1995), Nokia 3210 (1999), and Nokia 7110 (1999).

In the late 1990s, the enormous, and completely unexpected, uptake of SMS inspired attempts to bring the Internet to mobile handsets too. This led to the development of the Wireless Application Protocol (WAP) allowing simplified websites to be viewed on small displays and paving the way for Internet access on mobile devices. The first mobile phone to feature a WAP browser was the Nokia 7110. In response to the need for scrolling through long WAP pages it also featured the first "Navi-roller" thumb wheel. As an interesting example of interaction design, the 7110 also featured a *spring-loaded* cover concealing the keypad, which was inspired by the film The Matrix where the main character uses an earlier Nokia phone modified by the film's production crew to have this functionality. "Life imitating art" (Wilde 1889) you could say. WAP, however, never lived up to the expectations due to slow data transfer and poor usability (Ramsay and Nielsen 2000; Nielsen 2000) and was soon superseded by access to the real web on mobile devices. Nevertheless, mobile phone design in the 1990's had a fundamental and lasting impact on the future of mobile computing to come.

### 9.2.4 Convergence

One of the most interesting eras of mobile computing began when different types of specialised mobile devices began converging into new types of *hybrid* devices with fundamentally different form factors and interaction designs. The first phase of this was the emergence of "smart phones", which combined the functionality of a PDA with that of a mobile phone. The development of smart phones involved exploration of a wide range of form factors and interaction designs and led to a series of innovative solutions. Many of these involved designs where the physical shape of the device could be changed depending on what the user wanted to use it for. Other designs, like the Blackberry, introduced a "wide-body mobile phone" form factor with a PDA size display and a miniature QWERTY keyboard in place of the traditional 12-key numeric keypad. The first smart phone that, as well as making phone calls, could also be used for calendars, addresses, notes, e-mail, fax, and games was the IBM Simon from 1992. It had no physical buttons, but only a touch screen, which could be operated with a finger or a stylus.


NOKIA
Day
Month
Menu




ERICSSON
YES
NO
New
ink note
Online
services
Settings
R380e




FIGURE 9.6 A-B-C-D: Smartphones exploring different physical form factors and interaction styles: IBM Simon (1992), Nokia 9000 (1996), Ericsson R380 (2000), and Blackberry 5810 (2002).

The second phase of convergence combined mobile phones with various rich media capabilities, such as digital cameras, music players, video recording and playback, and television and radio reception. Whereas smart phones were attractive for business professionals' work activities and productivity, multimedia phones were attractive for everyday people's leisure, fun and socialising.


*Copyright © Sharp. All Rights Reserved. Used without permission under the Fair Use Doctrine (as permission could not be obtained). See the "Exceptions" section (and subsection "allRightsReserved-Used-WithoutPermission") on the page copyright notice.*


*Copyright © Nokia. All Rights Reserved. Used without permission under the Fair Use Doctrine (as permission could not be obtained). See the "Exceptions" section (and subsection "allRightsReserved-Used-WithoutPermission") on the page copyright notice.*







FIGURE 9.7 A-B-C-D: Converged mobile devices: camera-phones, game-phone and walkman-phone: Sharp J-SH04 (2001), Nokia N-Gage (2003), Nokia N90 (2005), and Sony Ericsson W600 (2005).

The most notable example of convergence for leisure was the invention of the camera phone. The first mobile phone to feature a digital camera was the Sharp J-SH04 from 2001. It was only available in Japan through the i-mode mobile Internet service, but the rest of the world soon followed. Two years later, more camera phones were sold than digital cameras, and in 2006 half the world's mobile phones had a built-in camera — making Nokia the biggest brand of digital cameras and forcing prominent brands such as Minolta and Konica out of the camera business. By 2009, there were more than 1.9 billion camera phones in existence, and mobile phone photography had already had a huge social impact through new ways of capturing and sharing photographs over the Internet (cf. Kindberg et al. 2005; Gye 2007). Whereas early camera phones were clearly phones with cameras, novel interaction design led to several converged devices truly blurring the boundaries between the two (Murphy et al. 2005). As an example, it can be hard to tell if the Nokia N90 is a phone or a camcorder. Another converged functionality to become widely available on mobile phones was the ability to listen to digital music. Most notably, Sony re-launched its successful "Walkman" brand of the 1980s in the shape of the converged Sony Ericsson W600 in 2005. With the W44 multimedia phone from 2006, they even went a step further and extended video and music playback with the ability to watch and listen to digital TV and radio. Convergence also led to the creation of hybrid game-phones like the Nokia N-Gage with form factors resembling handheld game consoles.

The fundamental driver behind the trend of convergence is that mobile user experience is proportionally related to the functional scope of interactive mobile devices and systems: "more means more" (Murphy et al. 2005). As a consequence, convergence has often been criticised for generating weak general solutions with usability comparable to the Swiss army knife: clumsy technology with a wide range of functions, none of which are ideal in isolation (see e.g. Norman 1998, Bergman 2000, Buxton 2001). However, in my view the real strength of convergence should not be sought in the simple availability of several functions implemented in the same device. Rather it should be found in the potential creation of something new and hybrid that facilitates use that wasn't possible before, like for example taking pictures and sharing them immediately with your friends, browsing the Internet on your phone, or purchasing music directly on your iPod.

### 9.2.5 Divergence

Contrasting the convergence approach, the trend of *divergence* suggested a single function/many devices or "information appliance" approach where each device is "designed to perform a specific activity, such as music, photography, or writing" (Bergman 2000). The driving force behind this line of thought is that having a wide range of good specialised tools is better than a general one that does not perform any task particularly well. Specialised tools facilitate optimisation of functionality over time and refinement of well-known paradigms of use. The fundamental view promoted by the trend of divergence is that mobile user experience is *inversely* proportionate to the functional scope of interactive mobile devices and systems: "less is more" (Murphy et al. 2005).


ARCHOS




POSITION
12/12
LAP RECORD
TOTAL TIME
LAP TIME
SONY
POWER
HOLD
HOME
VOL
SELECT
START




Figure 9.8 A-B-C-D: Specialised mobile media and gaming devices: Apple iPod (2001), Archos Gmini (2004), Sony PSP (2004), iPod Nano (2010).

The 2000s saw the emergence a wide range of diverged mobile devices dedicated to do one specific task really well, particularly mobile music players, video players and games. Of course functionally dedicated mobile devices were not a new phenomenon as, for example, early mobile devices such as pocket calculators, cell phones, GPS receivers, digital cameras, and PDAs could unarguably be classified as information appliances too. But what was interesting about the trend of divergence in the early 2000s was that it was a deliberate interaction design choice and not a technological necessity. Probably the most legendary example of an information appliance was the Apple iPod from 2001. Although not the first mobile digital music player, its interaction design, including the integration with iTunes and later the iTunes Music Store, fundamentally changed global music consumption and purchasing behaviour. Although most mobile phones on the market in the mid-2000s were able to play MP3 files, people still preferred to carry an additional device, the iPod, for playing their music as it provided a better user experience for that particular task, and the device itself had become a popular fashion item. In late 2010, the total number of iPods sold had exceeded 290 million units. Other diverged mobile devices included video players like the Archos Gmini from 2004, the Sony PSP game and video console, and later versions of the iPod extended with video playback capability, but within the same basic information appliance interaction design.

The interaction design challenge of a diverged mobile device is considerably different from that of a converged one because its functional scope is much narrower. However, as diverged devices are by definition typically used in concert with a plethora of other interactive devices and systems unknown to the designer, there is a huge interaction design challenge in supporting good and flexible integration and "convergence-in-use" (Murphy et al. 2005).

### 9.2.6 Apps

In June 2007, Apple launched the iPhone. Like many of its contemporaries this was a converged mobile device functioning as a camera phone, portable media player, and Internet client with e-mail, web browsing, and high-speed wireless network connectivity. However, rather than being just another incremental step in the evolution of converged mobile devices, the iPhone represented a significant rethinking of the design of mobile interactions and a series of notable interaction design choices. It featured a large high-resolution capacitive multi-touch display with simple gesture capabilities, such as swiping and pinching, and departed completely from the predominant use of physical keys and a stylus for text entry and interaction. Instead of navigating large and deep hierarchies of menus, the user experience was much more fluid and aesthetic, and the phone was both extremely easy and pleasurable to use. The iPhone also featured a number of embedded context sensors, which changed the orientation mode of the display depending on how it was held, as originally proposed in a UIST conference paper by Hinckley et al. (2000), and it thereby changed the mode of the phone application when held close to the face during a call. The later inclusion of GPS and a digital compass extended this "context-awareness" capability to also enable location-based services.

On the software side, the iPhone's web browser actually made it possible to access web content on a mobile device with a positive user experience, and many soon described handling email on the iPhone as favourable compared to its desktop counterparts. Dedicated applications provided direct access to watching video content from YouTube and purchasing music from the iTunes Store. In concert, this meant that people actually started using their mobile device as a *preferred* gateway to the Internet, rather than as a last resort. Consequently, iPhone OS dominated the total amount of mobile web traffic worldwide by mid-2009 (Admob 2009). In addition to this, data and media content can be integrated seamlessly with the user's other devices and computers at home or at work through cloud computing services such as MobileMe in a way never seen before in mobile interaction design, illustrating initial steps towards the creation of *digital ecosystems* of mobile and stationary computer systems connected through the Internet.

The iPhone completely redefined mobile computing and set new standards for mobile interaction design and user experiences that other companies, such as Google and HTC, still struggled to match up to 4 years later with the Android open source mobile operating system and associated online application store. In many ways, the iPhone was the device that mobile interaction design researchers had envisioned for a decade, and its enormous uptake worldwide, with over 120 million iOS enabled devices sold by September 2010, confirms that we were indeed right in our speculations about what people would want to do with mobiles — if only we could provide them with a good enough interaction design and user experience. The biggest impact of the iPhone, however, was not only in the interaction design of the device itself and in the high quality of its native applications. As it turned out, it was in the creation of an interaction design that provided users with easy access to a vast and unprecedented amount of *applications* for their mobile device.

In 2008, Apple launched the online “App Store” which provided a mechanism by which iPhone users could easily download, and pay for, third-party application content directly from their mobile device. These Apps span a range of functionalities, including social networking, productivity tools, personal utilities, games, navigation, and advertising for movies and TV shows. For creating this application content, an iPhone OS software development kit (SDK) was released for free along with a business model where Apple handles payments and distribution while leaving App creators with 70% of the profit. By 2012, more than 25 billion Apps had been downloaded from a selection of more than 500.000, making this hugely profitable for both Apple and for the individual third-party creators of particularly popular Apps, which in return has motivated the creation of even more application content. As an indication of the incredible size of this business, third party mobile software developers generated a total income of $2 billion by selling their products through the Apple App store in less than three years. Contrary to developing mobile applications in Java 2 Platform, Micro Edition (J2ME) or Qualcomm’s Binary Runtime Environment for Wireless (BREW), developing in iPhone SDK involves no need for customizing applications for a vast range of different handsets, which means that more time can be spent on application design. Also, in sharp contrast to the generally horrific mobile phone user interfaces for installing especially J2ME software, the iPhone provides not only a supply chain and billing model out-of-the-box, but also an application shopping user experience that is *positive in itself*. Hence, prior to the iPhone, downloading and installing software onto a mobile phone or PDA was something only technology-savvy people would do. Today this is common practice for millions of users, no matter their age and computing experience.

As an interesting effect of the iPhone-approach to mobile interaction design, improving the hardware specification of devices was suddenly surpassed in importance in favour of improving the software that is available for them. This is evidenced in the pace and scope of software developments and updates compared to equivalent hardware ones, which is an important shift within the design of mobile interactions. It indicates that a level of stability has been reached in terms of physical form factors and basic input and output capabilities, in favour of a focus on *applications and content*.

*Copyright © Apple Computer, Inc.. All Rights Reserved. Used without permission under the Fair Use Doctrine (as permission could not be obtained). See the "Exceptions" section (and subsection "allRightsReserved-UsedWithoutPermission") on the page copyright notice.*




Figure 9.9 A-B: The Apple iPhone and iPad (2007 and 2010).

Apple's success with the iPhone led to a third endeavour within mobile computing, the iPad, which was released in April 2010. Initial media reaction was mixed, but commercial uptake was unprecedented, and the iPad was sold in over 2 million units in its first two months, reaching 15 million units sold by the end of the year. While Microsoft's explicit interaction design approach for PDAs and tablets had long been to replicate the Windows 95 OS (Zuberec 2000), Apple took the opposite approach with the iPad tablet and based it on iPhone OS rather than MacOSX. This was a surprising move for many, admittedly including myself, but it had the effect of reinterpreting, and subsequently redefining, the so-far troubled category of "tablet computers" into a new category of *mobile devices* that are not just laptops without keyboards. Although criticised for being a closed system, the strength of the iPad lay in the user experience created through its meticulous interaction design, which invited the already growing community of iPhone interaction designers and application developers to explore the tablet form factor. Until then, nobody had cared to create web or native application content for tablets (Chen 2010), but with the iPad, tablets suddenly became one of the most interesting and promising mobile platforms on Earth, and by March 2011 there were more than 65.000 applications available for the iPad.

### 9.2.7 Digital ecosystems

As we move into the second decade of the new millennium, the challenges facing mobile computing and interaction design continue to evolve. The technical capabilities of our mobile devices have improved significantly to the point where factors such as screen real estate, input capabilities, processing power, network speed, and battery lifetime are much less of an issue than only half a decade ago. At the same time, we have also become sufficiently skilled at designing for relatively small screens and for the different input capabilities of mobile devices so that millions of ordinary people are actually able to download and use the applications being developed, and are even willing to pay for some of them. To a large extent, therefore, we have now successfully solved the majority of problems facing mobile interaction researchers and designers in the past. However, as the history of all areas of computing have shown us, it is highly unlikely that we have reached an end point. As in the past, the technology and interaction design we are witnessing today is just the starting point for the continuing evolution of the technology and interaction design of tomorrow. But what are then the challenges and opportunities for the design of mobile interactions to come? What will the next wave of mobile computing be about?

Fuelled by the enormous interest and uptake of “post-PC devices” like smart phones and tablets by the general population, it is not unreasonable to speculate that a major platform shift away from desktop computing is imminent. Mobile devices are becoming more and more important and widespread. They will soon be the dominating point of access to the Internet, and in combination with the growth of cloud computing they will soon dominate peoples' use of computational power. Importantly, what we are witnessing here is not just the development of even smarter smart phones with improved abilities to imitate desktop PCs in miniature. It is a radical evolution of a major computing platform for new applications allowing us to do things that couldn't be done before. This may well be a genuine paradigm shift for mobile computing and mobile interaction design.

Looking on the current trends, it appears that the next wave of mobile computing and interaction design is going to be about the creation of *digital ecosystems* (Miller et al. 2010) in which mobile computing plays a central role in concert with other ubiquitous computing resources. This challenges us to move beyond considering interactive mobile devices, systems, and services as entities that can meaningfully be designed and studied in isolation from the larger use context or artefact ecologies (Jung et al. 2008, Bødker and Klokmose 2011) that they are a part of. Yes, mobile computers, in various forms, play hugely important roles in most peoples' everyday lives, but they are not the only technologies and artefacts we make use of at home or at work, or in the space between. Most people use multiple mobile devices for different purposes, but they also use a multitude of stationary or embedded computer systems, at work, at home, in their cars, or in the city around them. In concert, this makes up a rich digital ecosystem of interactive devices, systems and services often referred to as ubiquitous or pervasive computing, in which mobile computing is a central, but not the only, component. The challenge of designing mobile interactions in such ubiquitous and pervasive information societies is to facilitate the creation of interactive devices, systems, and services that fit well into this ecosystem of other devices, systems, and services, as well as into the rich new use patterns, for work and leisure, created by these technologies and their users. Like any other type of ecosystem, understanding, creating, and maintaining digital ecosystems requires a holistic perspective on the totality and ecology of the system at play, and not just detailed views on each of its individual components. The digital ecology wave of mobile computing will build on the achievements of previous eras within hardware miniaturization, connectivity, new form factors, input devices, interaction styles, applications, convergence, divergence, and content, but it will broaden the scope to include the wider context of use and an explicit sensitivity for the contextual factors that influence the user experience. It is going to be about creating interactive devices, systems, and services that respond to the broad and diverse aspects of human life, and these not only provide utility and are easy to use, but also provide pleasure and fit naturally into peoples' complex and dynamic lives of constantly changing settings and situations.

## 9.3 INTERACTION DESIGN FOR MOBILE COMPUTERS

The term interaction design, coined by Bill Moggridge and Bill Verplank in the late 1980s, is about "designing interactive products to support the way people communicate and interact in their everyday and working lives" (Sharp et al. 2007 p. 8), or more broadly about "the design of everything that is both digital and interactive" (Moggridge 2007 p. 660) with particular attention to its subjective and qualitative aspects. In other words, it is about creating life and work enhancing user experiences through the design, development, construction, and implementation of interactive products, devices, systems, and services.

Today, interactive products are typically computer-based, and this means that interaction design is relevant within all disciplines, fields, and approaches that concern themselves with research and design of computer-based systems for people. Hence, alongside design practices such as graphic and industrial design, academic disciplines such as psychology and sociology, and multi/interdisciplinary fields such as human-computer interaction and information systems, interaction design also involves the technical academic disciplines of computer science and engineering. However, interaction design differs from each of these practices, disciplines, and fields by having a different, overall, focus and purpose. It is concerned with the *totality* of the user experience of interactive products and with all of the factors that may contribute to their successful creation. When we design computer-based interactive systems, we are not just designing how it appears, but also how it behaves. We are designing how people and technology interact (Moggridge 2007). As described by Winograd (1996), doing interaction design can in many ways be compared to doing architecture. The architect is concerned with people and their interactions within the building being created. For example, does the space fit the lives or work styles of the family or business that is going to inhabit it? Does the flow within and between rooms work well? Are functionally related spaces in close proximity? And so on. Supporting the work of the architect, engineers are concerned with the structural soundness and construction methods of the building, and knowledge from other disciplines, such as human factors and social sciences, may also influence the architect's ability to create functional and liveable spaces. Just like a good architect understands these other relevant disciplines, so does a good interaction designer. However, just like there is a difference between designing and building a house there is also a difference between designing an interactive product and engineering its software (Sharp et al. 2007 p. 9).

Mobile interaction design is an area of interaction design that is concerned specifically with the creation of user experiences with interactive products, devices, systems, and services that are not stationary, but that people can take with them. It is enabled by advances in mobile computing — as described earlier — that have allowed designers and system developers to conceive interactive products that are small enough to be carried with us, held in our hands, or even worn, while also providing computational power and network capabilities sufficient enough for enabling useful and attractive interactive systems and services. This includes handheld and wearable devices, PDAs, mobile phones, smart phones, portable digital media players, handheld games, etc. as well as the software applications and services that run on these devices or can be accessed from them. However, mobile interaction design is not only facilitated and driven by advances in computer science and engineering. It is also increasingly advanced by our ability to develop new use practices for mobile computing and to include and appropriate available and emerging mobile computer and network technologies into new and innovative interactive products and solutions. Hence, we have long ago gone beyond the "anytime anywhere" mobile computing hype of the late 1990s and grown much more sensible aspirations to develop "mobiles that work at the right time, and that know their place — that fit in" (Jones and Marsden 2006).

The challenges of mobile interaction design have changed and evolved over time as new technologies were developed and new use practices emerged. Early mobile interaction design dealt with the physical design of portable computers. This evolved into a focus on input devices and interaction styles suitable for handheld operation and mobile use. For mobile phones, the interaction design challenge has primarily been a matter of reducing physical size while optimizing the use of limited display real estate and the standard 12-key numeric keypad for more and more possible applications. With the emergence of functionally hybrid and more complex devices, the interaction design challenge became about developing new forms and shapes of devices as well as developing new types of applications available on them, without making the devices (even) harder to use. For the growing range of functionally dedicated mobile devices like digital cameras and media players, the interaction design challenge became about facilitating peoples' "orchestration" of all these devices, and their content, in increasingly complex ecosystems of interactive computer systems and digital data.

Today, the challenge of designing mobile interactions is very much about the development of *software applications*. The physical device form factor appears to have stabilized, for some time at least, on the basic size, shape, and interaction capability introduced by the Apple iPhone in 2007, which has remained unchanged for more than four years and been replicated by all major handset producers. This has shifted focus towards downloadable and purchasable third party application content available for these devices, in the form of relatively small “Apps” with highly specialised functionality, designed not only by large software corporations, but also by small companies and even individuals, including students. By late 2010, more than 300.000 third party applications were available from the Apple App Store, and more than 80.000 were available from Google’s Android Market. In less than three years, more than 10 billion Apps were downloaded for the iPhone and iPod Touch. However, although a lot of interesting and innovative new mobile applications are appearing in Google’s and Apple’s online stores every day, and application developers and interaction designers worldwide are pushing the boundaries of what mobile computer devices are being used for, the state of current mobile application design can be compared to the state of the web in the mid-1990s. There is a lot of excitement and interest, the development tools are easily accessible, and there is a huge audience of potential users. Exceeding the potentials of the web in the mid- 1990s, there are even well established digital supply chains and mechanisms for micro-payments. But as with the web 15 years ago, we haven’t yet seen or understood the significance and scope of the impact that third party application design for mobile devices will have on all aspects of our lives, for work as well as for leisure.

### 9.3.1 The role of context

Since the early days of mobile computing and mobile human-computer interaction, the use contexts of interactive mobile systems and devices have often been highlighted as being particularly important for system developers to “be aware of” and “take into account” when designing and building interactive mobile systems, and when evaluating and studying their use (cf. Johnson 1998, Rodden et al. 1998, Brown et al. 2000). Mobile use contexts have been described as being particularly challenging compared to, for example, the use contexts of traditional stationary office systems due to their highly dynamic, complex, and indeed mobile, nature. It has also often been suggested that when using an interactive mobile computer system, other activities in the surrounding context are often more important than the actual interaction with and use of the system itself — walking down the street, socialising in a bar or café, or attending to a patient in a hospital.


*Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.*


*Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.*


*Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.*

FIGURE 9.10 A-B-C: Mobile Computing in context (Kjeldskov and Paay 2010).

There are many different definitions of context, and the debate on what constitutes context for mobile computing, and what role it plays, is ongoing. Early works within mobile computing referred to context as primarily the location of people and objects (Schilit and Theimer 1994). In more recent works, context has been extended to include a broader collection of factors such as physical and social aspects of an environment (McCullough 2004, Dourish 2004, Bradley and Dunlop 2002, Agre 2001, Dey 2001, Abowd and Mynatt 2000, Schmidt et al. 1999a, Crabtree and Rhodes 1998). Dey (2001) defines context as “any information that can be used to characterise the situation of an entity. An entity is a person, place or object that is considered relevant to the interaction between a user and an application, including the user and the application themselves.” Although this definition is quite complete, it is not very specific about what type of information could in fact be used to characterise such a situation. In contrast to this, Schmidt et al. (1999a) present a model of context with two distinct categories: human factors and physical environment. Human factors consist of the three categories: information about the user (profile, emotional state, etc.), the user’s social environment (presence of other people, group dynamics, etc.), and the user’s tasks (current activity, goals, etc.). Physical environment consists of the three categories: location (absolute and relative position, etc.), infrastructure (computational resources, etc.), and physical conditions (noise, light, etc.). This model provides a good catalogue of specific contextual factors to complement broader definitions like the one by Dey (2001). Other works are not as comprehensive in their coverage of different contextual factors, but go into detail about one or a few. In the works of Agre (2001) and McCullough (2004), particular importance is given to physical context consisting of architectural structures and elements of the built environment, for example, landmarks and pathways. In the works of Dourish (Dourish 2001, Dourish 2004), particular importance is given to social context including interaction with, and the behaviour of, people in an environment. Dourish (2004) also states that context cannot be defined as a stable description of a setting, but instead arises from, and is sustained by, the activities of people. Hence, it is continually being renegotiated and redefined in the course of action. These works provide us with additional contextual factors of particular relevance to mobile computing in context, and with the knowledge that what defines context is in itself contextually dependent.

The context of mobile computing is something that several individual disciplines within mobile interaction design are concerned with, and that has influenced the shaping of methodology, technology, and theory within and across the field's internal disciplinary boundaries. These different disciplines have each approached the challenge of contexts differently, and have yielded different types of responses.

In domain studies of mobile computing, where context plays an obvious central role as essentially the phenomenon under scrutiny, the challenge has been partly to understand theoretically what use contexts are and how they can be described, and partly to study empirically what characterises specific use contexts of interest, and how the phenomenon of context can be studied and analysed in ways that generate such understanding. This has led to a body of theoretical and socio-technical research building largely on methods and theories from sociology, anthropology, and phenomenology (e.g. Luff and Heath 1998, Dourish 2001, Dourish 2004, Dey 2001, Ling 2001, Perry et al. 2001, Fortunati 2001, Green et al. 2001, Agre 2001, McCullough 2004, Chalmers 2004, Aoki et al. 2009, Kostakos et al. 2009), as well as my own work in this area (Paay and Kjeldskov 2005, Paay and Kjeldskov 2008a, Kjeldskov et al. 2004, Kjeldskov and Stage 2006).

In systems development and design for mobile computing, the challenge of context has primarily been about creating an appropriate fit between systems and context and how this can be supported structurally through new, or modified, systems development and design methods. While relatively very little has been published on this topic, there is an emerging body of methodological research building largely on methods and theories from information systems, software engineering and human-computer interaction (e.g. Sharples et al. 2002, Mikkonen et al. 2002, Hosbond 2005, Paay 2008, de Sá and Carrico 2009, Paay et al. 2009a), as well as my own work (Kjeldskov and Howard 2004, Paay et al. 2009b, Vetere et al. 2005, Kjeldskov and Stage 2012).

In usability evaluation for mobile computing, the challenge of context has primarily been to understand its role in relation to the scope, richness, and validity of empirical findings and how usability tests can be carried out in contextually realistic settings through use of new or modified methods and techniques. This has led to a growing body of empirical research building largely on methods and theories from usability engineering. These include, for example (Brewster 2002, Betiol and Cybis 2005, Hagen et al. 2005, Kaikkonen et al. 2005, Nielsen et al. 2006, Rogers et al. 2007, Reichl et al. 2007, Oulasvirta 2009, Oulasvirta and Nyyssonen 2009, de Sá and Carrico 2010), as well as my own contributions (Kjeldskov and Stage 2004, Kjeldskov et al. 2004, Kjeldskov et al. 2005, Kjeldskov and Skov 2007a, Høegh et al. 2008).

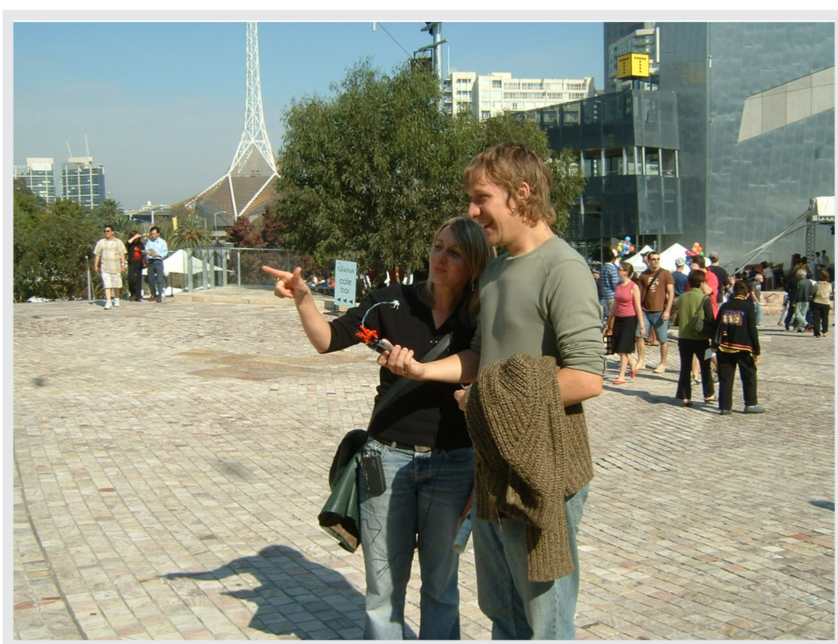
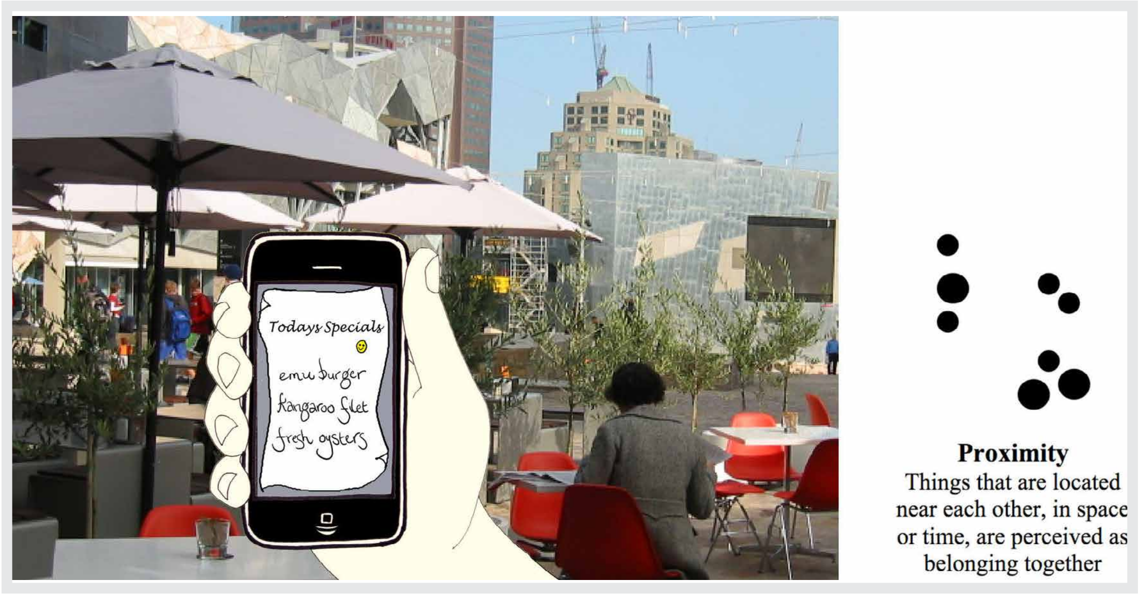


Figure 9.11: Evaluating mobile computing in context.

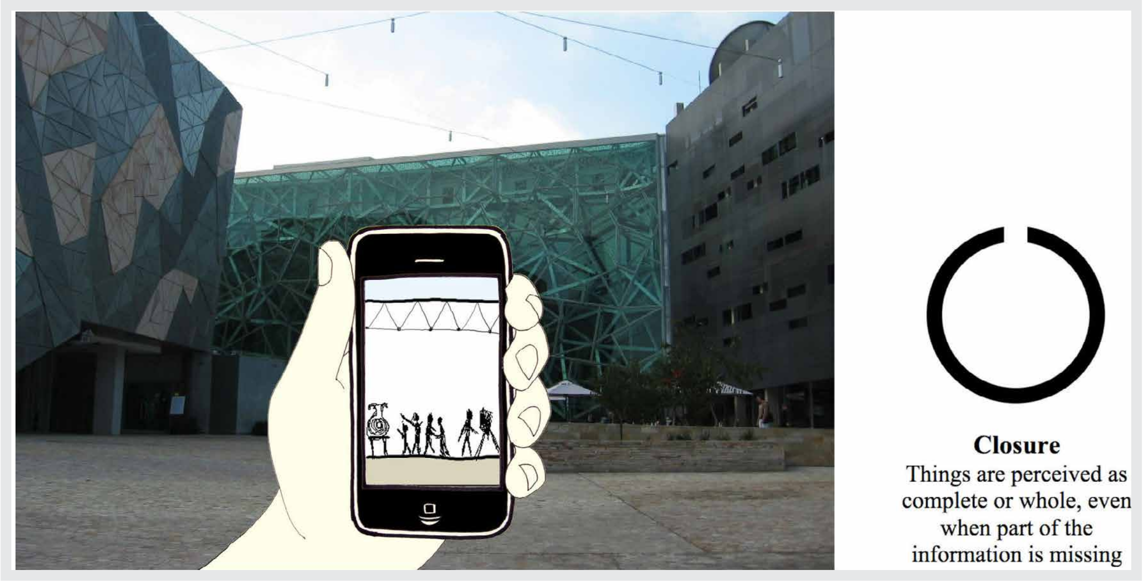
*Copyright:*

In implementation of mobile computing, the challenge of context has largely been about capturing, formalizing, and modelling this attribute in computational data models, how to make sense from such models, and how to use them in the construction of *context-aware* mobile systems that are responsive to their surroundings. This has led to an extensive body of technical research building largely on methods and theories from computer science (e.g. Schilit and Theimer 1994, Crabtree and Rhodes 1998, Schmidt et al. 1999a, Schimdt et al. 1999b, Cheverst et al. 2001, Dix et al. 2000, Chen and Kotz 2000, Hinckley and Horvitz 2001, Dey 2001, Jameson 2001, Jones et al. 2004, Edwards 2005, Hinckley et al. 2005), as well as my own contributions (Kjeldskov and Skov 2007b, Kjeldskov and Paay 2005, Kjeldskov and Paay 2006, Kjeldskov et al. 2010, Skov et al. 2012, Kjeldskov et al. 2012).

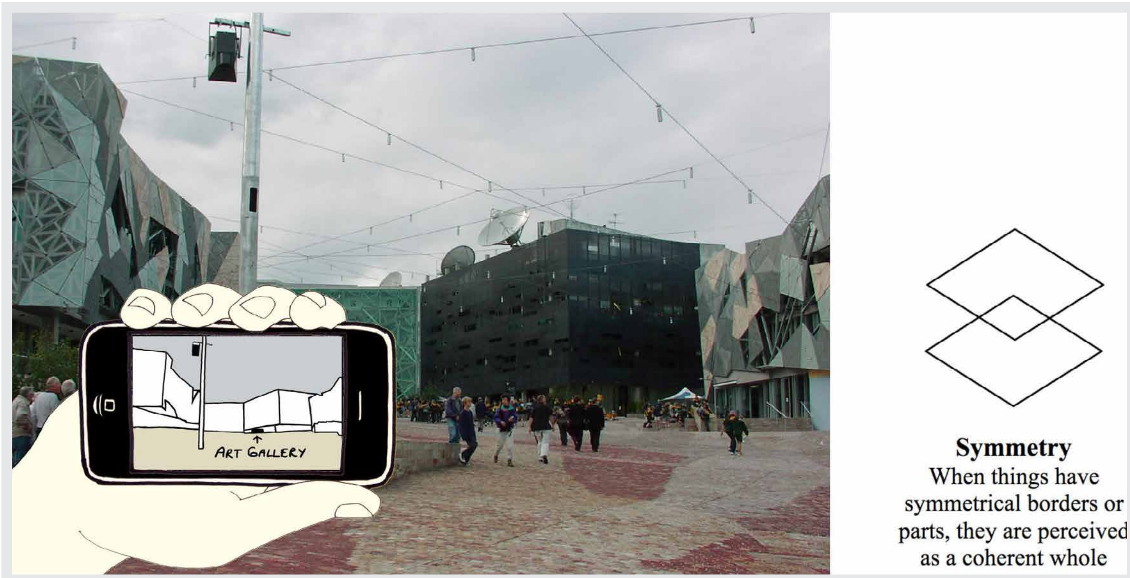
In user experience research for mobile computing, the challenge of context has been to understand what impact rich and dynamic user contexts have on peoples' experience of using technology, and to describe how this user experience can be improved. This has led to a body of theoretical, conceptual, and design-oriented research building on methods and theories from a wide range of disciplines from sociology and psychology to cognitive science, computer science, human-computer interaction, and computer-supported cooperative work. These include, for example (Abowd and Mynatt 2000, Cheverst et al. 2001, Palen et al. 2000, Weilenmann 2001, Bradley and Dunlop 2002, Brown and Randell 2004, Little and Briggs 2009, Benford et al. 2009, Karapanos et al. 2009, Lindley et al. 2009, Rowland et al. 2009), as well as my own contributions (Paay and Kjeldskov 2008b, Kjeldskov and Paay 2010, O'Hara et al. 2011, Murphy et al. 2005).



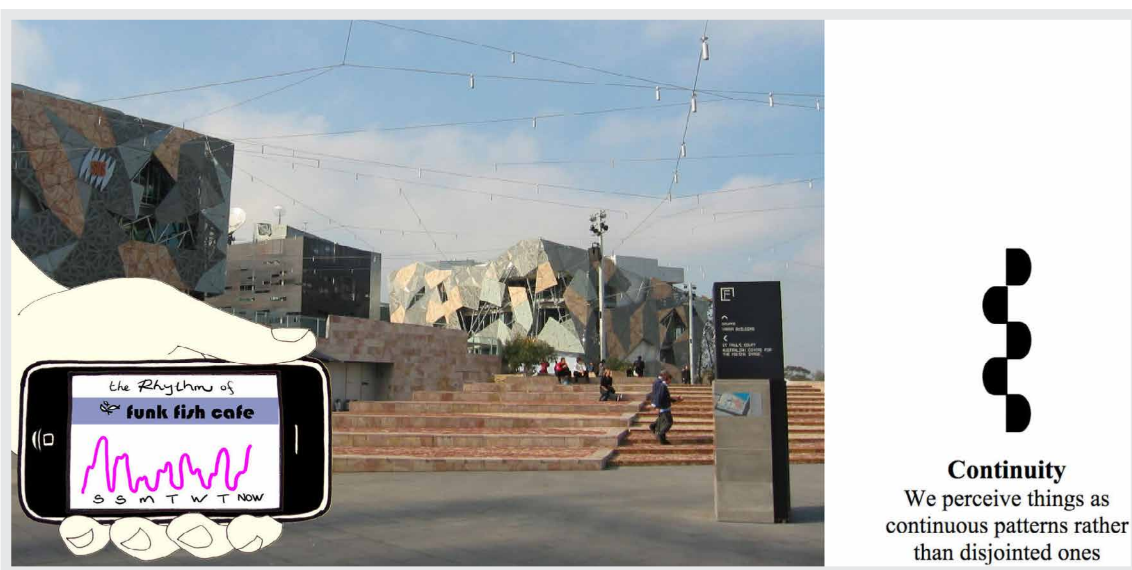

*Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.*




*Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.*


ART GALLERY
Symmetry
When things have symmetrical borders or parts, they are perceived as a coherent whole




the Rhythm of
funk fish cafe
S S M T W T NOW
Continuity
We perceive things as continuous patterns rather than disjointed ones

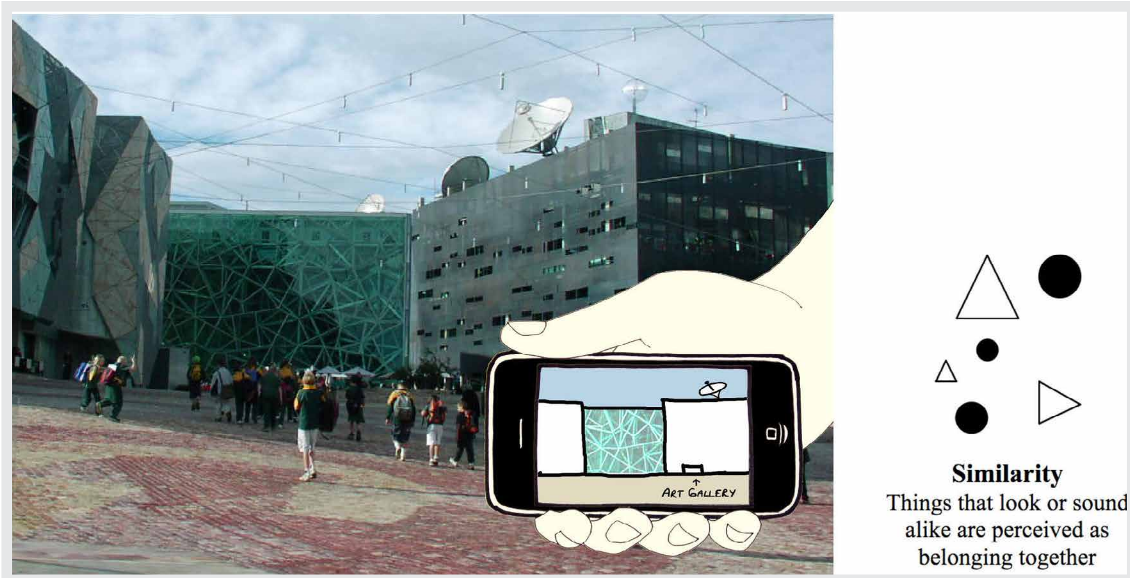





Figure 9.12: Explaining mobile user experience in context using five principles of perceptual organisation from Gestalt Theory (Paay and Kjeldskov 2008b).

This is not to say that context is a new phenomenon appearing on the research agenda with the emergence of *mobile* computing. Context has indeed been an important concept within human-computer interaction and interaction design since the *second wave* or *paradigm* of HCI (Bødker 2006, Harrison et al. 2007). The first wave of HCI was a mixture of engineering and human factors focussing on optimizing human-machine fit. The second wave was largely based on cognitive science focussing on the simultaneous processing of information in machines and in the human mind, but this also involved a strong focus on the use of interactive computing systems in the context of the workplace. However, as pointed out by Bødker (2006), while there was lot of discussion about the intricate concept of context in second wave HCI, this research achieved little in terms of defining and operationalising it in a way of any real significant value to HCI and interaction design. In the *third* wave, focus has broadened further towards a post PC ubiquitous and pervasive information society where computer technology has spread "from the workplace to our homes and everyday lives and culture" (Bødker 2006). This means that context is now an *elemental* concept that we not only need to *define* well, but also need to *understand* better in terms of its complexity, significance, and influence on peoples' experience of technology in use, in order to inform technology design better.

Mobile interaction design is positioned within the second and third waves of HCI. It grew out of the second wave, but the tremendous uptake of mobile computing by the general population subsequently was a contributing factor to the creation, force, and velocity of the third wave by enabling some of the completely new potentials and patterns of computing technology use that we are witnessing globally today.

### 9.3.2 Research impact on practice

Much of the future impact of mobile computing envisioned earlier will be driven by skilful and creative design of mobile interactions conceived by entrepreneurial developers and designers who understand how to create useful and enjoyable utility and user experience that fits the user's needs, desires and contexts of use. Unfortunately, however, the current research-based literature on mobile interaction design neither provide as much foundation as we probably could for these developers and designers to base their innovations and interaction design on, nor much methodological guidance on how to approach the process. Whereas there are a lot of research-based books about user interface and interaction design for desktop applications and web sites, there is not yet a lot of equivalent literature available about mobile interaction design. Although mobile computing has a history of approximately three decades, and interaction design has played an important role throughout about 2/3 of this history, only one good general textbook, by Jones and Marsden (2006), has been published on the topic to date. And although this book is indeed a brilliant starting point for addressing the particular challenges of *mobile* interaction design, it still doesn't have the completeness and depth of equivalent human-computer interaction and interaction design primers such as Laurel (1990), Shneiderman (1997), Preece et al. (1994), Winograd (1996), Raskin (2000), Dix et al. (2004), Benyon et al. (2005), Lauesen (2005), Bagnara and Smith (2006), Preece et al. (2002), and Rogers et al. (2011). This is potentially an opportunity missed for large-scale real-world impact on mobile interaction design *practice* in respect to the massive amount of good interaction design research that has been done within the field over the last decade and a half. While it might indicate that the area of mobile interaction design still hasn't stabilized enough for general guidelines, principles, methods, and techniques to evolve, it also demonstrates an opportunity, and a need, to push forward on developing such foundational work further.

Several of the textbooks that *do* exist on aspects of interaction design for mobile devices, systems, and services, such as Helal et al. (1998), Weiss (2002), Ballard (2007), Fling (2009) and Frederick and Lal (2010), essentially target application development for particular and very specific classes of devices and software platforms, and address ephemeral technical limitations, such as particular operating systems, low screen resolution, reduced processing power, limited memory, and poor bandwidth. While unarguably useful when designing for these exact platforms, the weakness of such types of works is that they are almost *too* practical. They are highly vulnerable to technological advances and therefore quickly rendered irrelevant as new devices and platforms emerge. As a consequence, they usually end up as short-lived and overly specific user interface guidelines tied to a specific point in time, and not as generally applicable and timeless principles for interaction design. Distilling the essence of these works — the higher-level challenges and solutions that apply beyond specific devices and platforms — would be useful for moving the field of mobile interaction design forward. But such work has not yet been done systematically and in depth.

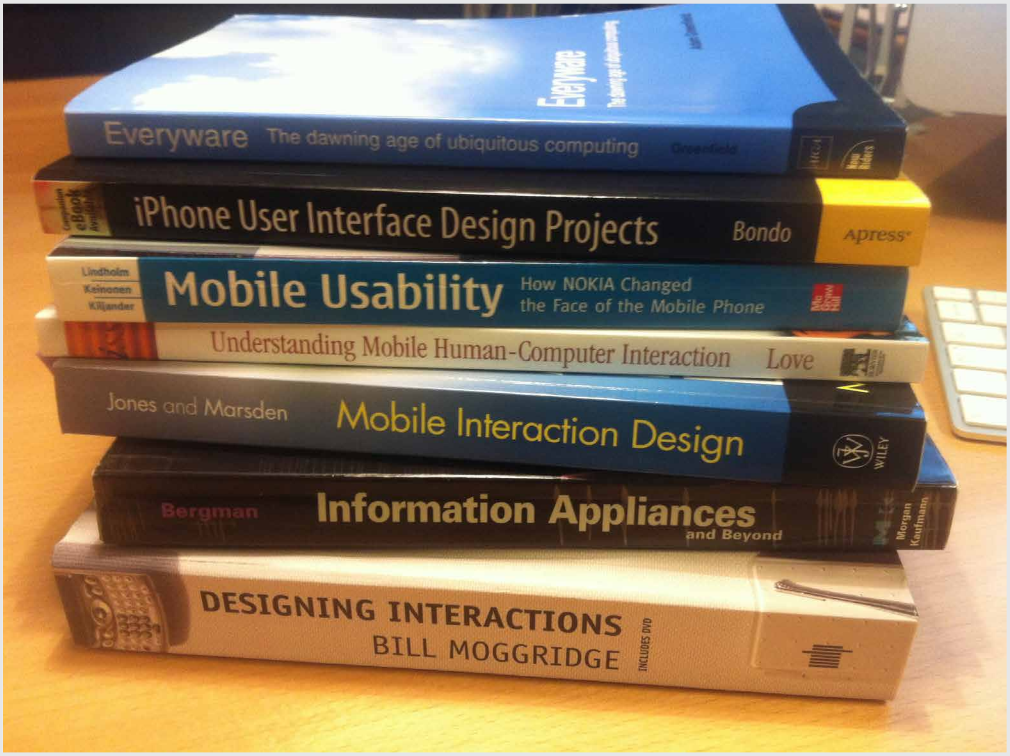



Figure 9.13: Some suggested reading.



As a step in the right direction, though, a different class of textbooks on mobile interaction design is the collection of case study-like accounts for successful and influential design solutions, such as Eric Bergman's "Information Appliances and Beyond" (Bergman 2000), Lindholm and Keinonen's "Mobile Usability: how Nokia changed the face of the mobile phone" (Lindholm and Keinonen 2003), parts of Bill Moggridge's "Designing Interactions" (Moggridge 2007), and Bondo et al's "iPhone User Interface Design Projects" (Bondo et al. 2009). These writings aim to capture universally important lessons learned from the experience of actual mobile interaction designers. They provide interaction design as well as methodological insight about influential solutions and how they came about. The potential weakness of *these* works, however, is that they easily end up being anecdotal and difficult to transfer into present time's design challenges. To support such transfer and transcendence of knowledge, we must provide not only the case study accounts, but also analysis across these case studies that elevates our learning from the concrete and specific level to the abstract and general.

## 9.4 THE ROAD AHEAD: TOWARDS DIGITAL ECOLOGY

So where do we go from here? As I have discussed earlier, the currently emerging trend within mobile computing is the creation of digital ecosystems where interactive mobile systems and devices are viewed less in isolation and more as parts of larger use contexts or artefact ecologies (see, for example, Jung et al. 2008, Bødker and Klokmose 2011). From my perspective, this is an avenue for further research and design that is particularly interesting, and one that I personally look forward to engaging myself in more deeply.

As a starting point for this, I believe that we need to develop interaction design approaches that focus more broadly on "the whole" and have a build-in sensitivity for the continual emergence and convergence of systems and their context that characterises such ecosystems and ecologies. Contemporary interactive mobile systems, services, and devices have become integral parts of ubiquitous computing environments that we care deeply about. However, although their look, feel, and features impact our everyday lives as we orchestrate them in concert with a plethora of other computing technologies, these artefacts and ecosystems are not well understood or created through traditional methods of user-centred design and usability engineering. Contrary to more traditional IT artefacts, they constitute holistic user experiences of value and pleasure that require careful attention to the variety, complexity and dynamics of their usage. Hence, we need further development of theoretical and conceptual lenses through which we can view, address, and describe this emerging phenomenon in a way that informs and inspires design and further thinking. This work may find inspiration and traction in some of the conceptually stronger and less technical literature on ubiquitous and pervasive computing that has started to appear in recent years, such as Adam Greenfield's book "Everyware" (Greenfield 2006).

As a way of encapsulating and labeling this work, I suggest to use and develop the term *digital ecology*. Ecology is the study of elements making up an ecosystem, and is very generally about understanding the interactions between organisms and their environment. It is inherently holistic and has an interdisciplinary nature, and it is not synonymous with "the environment" or with "environmentalism". Nor is ecological thinking limited to the discipline of biology. For example, "industrial ecology" studies material and energy flows through networks of industrial processes, and "human ecology" is an interdisciplinary area of research that provides a framework for understanding and researching human social interaction. In a similar fashion, I believe "digital ecology" may be a useful way of describing the study of elements making up digital ecosystems and the holistic understanding of interactions between these elements and their environment. By digital ecology is thereby meant the study of interrelated digital systems (e.g. mobile and pervasive computing) and the processes by which these systems work and interact, and are conceived, emerge, converge, and evolve. It is about understanding the functioning, use and experience of digital ecosystems and artefact ecologies around us, and the design processes that create and advance them.

## 9.5 WHERE TO LEARN MORE

There are several online sources for more information. Some are available for free, while others require a subscription or accessing them from within subscribing universities' networks.

### 9.5.1 Conferences

The Mobile HCI conference series is a central place to go for more information. Proceedings from the conference series are available electronically through Springer and ACM. Mark Dunlop, one of the initiators of the conference series, keeps a general page on the conference at and there is also a page on Wikipedia http://en.wikipedia.org/wiki/MobileHCI worth visiting as a gateway.

In addition, proceedings from the ACM CHI conference series contain numerous articles on human-computer interaction with mobile computer systems. These proceedings can be accessed on http://dl.acm.org/

In addition to these, it's worth browsing the Proceedings from the UbiComp conference series, the MobiCom conference series, and the Pervasive conference series).

### 9.5.2 Journals

Many journals in HCI have published articles on mobile computing. Among the ones dedicated to the topic are:

- *Personal and Ubiquitous Computing*
- Pervasive and Mobile Computing
- *International Journal of Mobile Human Computer Interaction*

### 9.5.3 Recommended reading

From the references below, I particularly recommend these articles and books.

- Atkinson, P. (2005) Man in a Briefcase - The Social Construction of the Laptop Computer and the Emergence of a Type Form. *Journal of Design History*, 18(2), 191-205.
- Bergman E. (Ed.) (2000) *Information Appliances and Beyond*. San Francisco: Morgan Kaufmann Publishers.
- Greenfield, A. (2006) *Everyware: the dawning age of ubiquitous computing*. Berkeley: New Riders.

- Hinckley, K., Pierce, J., Sinclair, M. and Horvitz, E. (2000) Sensing techniques for mobile interaction. In *Proceedings of UIST 2000* (pp. 91-100). New York: ACM.
- Johnson, P. (1998) Usability and Mobility; Interactions on the move. In *Proceedings of the First Workshop on Human-Computer Interaction with Mobile Devices*, Glasgow, Scotland (GIST Technical Report G98-1).
- Jones, M. and Marsden, G. (2006) *Mobile Interaction Design*. Glasgow: John Wiley and Sons, Ltd.
- Kay, Alan (1972). A Personal Computer for Children of All Ages. In *Proceedings of ACM National Conference*. Boston: New York: ACM.
- Paay, J., (2008) From ethnography to interface design. In J. Lumsden (Ed.), *Handbook of Research on User Interface Design and Evaluation for Mobile Technology* (pp. 1-15). PA, USA: Idea Group Inc (IGI).
- Perry, M., O'Hara, K., Sellen, A., Brown, B. and Harper, R. (2001) Dealing with mobility: understanding access anytime, anywhere. *ACM Transactions on Computer-Human Interaction, 8*(4), 323-347.
- Rogers, Y., Connelly, K., Tedesco, L., Hazlewood, W., Kurtz, A., Hall, R. E., Hursey, J. and Toscos, T. (2007) Why it's worth the hassle: the value of in-situ studies when designing Ubicomp. In *Proceedings UbiComp 2007*, LNCS (pp. 336-353). Berlin: Springer-Verlag.
- Weiser, M. (1991) The Computer for the 21st Century. *Scientific American, 265*(3), 94-104.

### 9.5.4 Recommendations from my own publications on the topic

Available from people.cs.aau.dk/~jesper/


- Kjeldskov J. (2012) Designing mobile interactions — the continual convergence of form and context. Volume 1 and 2 (forthcoming)
- Kjeldskov, J., Skov, M. B., Nielsen, G. W., Thorup, S. and Vestergaard, M. (2012) Digital Urban Ambience: Mediating Context on Mobile Devices in the City. *Journal of Pervasive and Mobile Computing* (in press).
- Kjeldskov J., Cheverst K., de Sá M., Jones M., and Murray -Smith R. (2012) Research Methods in Mobile HCI: Trends and Opportunities. *Proceedings of Mobile HCI 2012 (vol. 2),* September 21-24, San Francisco, USA. ACM Press, pp. 255-260.
- Kjeldskov J. and Paay J (2012) A longitudinal review of Mobile HCI research Methods. *Proceedings of Mobile HCI 2012,* September 21-24, San Francisco, USA. ACM Press, pp. 69-78.
- Kjeldskov, J. and Paay, J. (2010) Indexicality: understanding mobile human-computer interaction in context. *ACM Transactions on Computer-Human Interaction (TOCHI). 17*(4)
- Kjeldskov, J. and Paay, J. (2006) Public Pervasive Computing in the City: Making the Invisible Visible. *IEEE Computer, 39*(9), 60-65.
- Kjeldskov, J. and Paay, J. (2005) Just-for-Us: A Context-Aware Mobile Information System Facilitating Sociality. In *Proceedings of Mobile HCI 2005*, Salzburg, Austria (pp. 23-30). New York: ACM.
- Kjeldskov, J., Graham, C., Pedell, S., Vetere, F., Howard, S., Balbo, S. and Davies, J. (2005) Evaluating the Usability of a Mobile

Guide: The influence of Location, Participants and Resources. *Behaviour and Information Technology, 24*(1), 51-65.

- Kjeldskov, J., Skov, M. B., Als, B. S. and Høegh, R. T. (2004a) Is it Worth the Hassle? Exploring the Added Value of Evaluating the Usability of Context-Aware Mobile Systems in the Field. In *Proceedings of MobileHCI 2004*, Glasgow, Scotland, LNCS (pp. 61-73). Berlin: Springer-Verlag.
- Kjeldskov, J. and Stage, J. (2004) New Techniques for Usability Evaluation of Mobile Systems. *International Journal of Human-Computer Studies, 60*(2004), 599-620.
- Kjeldskov, J. and Graham, C. (2003) A Review of MobileHCI Research Methods. In *Proceedings of the 5th International Mobile HCI 2003 conference*, Udine, Italy, LNCS (pp. 317-335). Berlin: Springer-Verlag.
- Paay, J., Kjeldskov, J., Howard S. and Dave, B. (2009) Out on the town: a socio-physical approach to the design of a context aware urban guide. *Transactions on Computer-Human Interaction, 16*(2), 7-34.
- Paay, J. and Kjeldskov, J. (2008) Understanding the user experience of location based services: five principles of perceptual organization applied. *Journal of Location-Based Services, 2*(4), 267-286

## 9.6 REFERENCES

*Abowd, Gregory D. and Mynatt, Elizabeth D. (2000): Charting Past, Present, and Future Research in Ubiquitous Computing. In ACM Transactions on Computer-Human Interaction, 7 (1) pp. 29-58*

*Agre, Philip E. (2001): Changing Places: Contexts of Awareness in Computing. In Human-Computer Interaction, 16 (2) pp. 177-192*

Aoki, Paul M., Honicky, R. J., Mainwaring, Alan, Myers, Chris, Paulos, Eric, Subramanian, Sushmita and Woodruff, Allison (2009): A vehicle for research: using street sweepers to explore the landscape of environmental community action. In: Proceedings of ACM CHI 2009 Conference on Human Factors in Computing Systems 2009. pp. 375-384

Atkinson, P. (2005): Man in a briefcase: the social construction of the laptop computer and the emergence of a type form. In Journal of Design History, 18 (2) pp. 191-205

Bagnara, Sebastiano and Smith, Gillian Crampton (2006): Theories and Practice in Interaction Design (Human Factors and Ergonomics Series). Lawrence Erlbaum Associates

Balakrishnan, Ravin, Fitzmaurice, George W., Kurtenbach, Gordon and Singh, Karan (1999): Exploring interactive curve and surface manipulation using a bend and twist sensitive input strip. In: SI3D 1999 1999. pp. 111-118

Ballard, Barbara (2007): Designing the Mobile User Experience. Wiley

Benford, Steve, Giannachi, Gabriella, Koleva, Boriana and Rodden, Tom (2009): From interaction to trajectories: designing coherent journeys through user experiences. In: Proceedings of ACM CHI 2009 Conference on Human Factors in Computing Systems 2009. pp. 709-718

Benyon, David, Turner, Phil and Turner, Susan (2005): Designing interactive systems : people, activities, contexts, technologies. Addison-Wesley

Bergman, Eric (2000): Information Appliances and Beyond: Interaction Design for Consumer Products (Interactive Technologies). Morgan Kaufmann

Bergman, Eric and Haitani, Rob (2000): Designing the PalmPilot: a Conversation with Rob Haitani. In: Bergman, Eric (ed.). "Information Appliances and Beyond: Interaction Design for Consumer Products (Interactive Technologies)". Morgan Kaufmann

Betiol, A. H. and Cybis, W. de Abreu (2005): Usability Testing of Mobile Devices: A Comparison of Three Approaches. In: Proceedings of IFIP INTERACT05: Human-Computer Interaction 2005. pp. 470-481

Bondo, Joachim, Barnard, David, Wilson, Eddie, Peters, Keith, Kemper, Craig, Burcaw, Dan, Novikoff, Tim and Parrish, Chris (2009): iPhone User Interface Design Projects. Apress

Bradley, Nicholas A. and Dunlop, Mark D. (2002): Understanding Contextual Interactions to Design Navigational Context-Aware Applications. In: Paterno, Fabio (ed.) Mobile Human-Computer Interaction - 4th International Symposium - Mobile HCI 2002 September 18-20, 2002, Pisa, Italy. pp. 349-353

*Brewster, Stephen A. (2002): Overcoming the Lack of Screen Space on Mobile Computers. In Personal and Ubiquitous Computing, 6 (3) pp. 188-205*

*Brown, Barry (2004): Building a Context Sensitive Telephone: Some Hopes and Pitfalls for Context Sensitive Computing. In Computer Supported Cooperative Work, 13 (3) pp. 329-345*

*Brown, Barry, Sellen, Abigail and O'Hara, Kenton P. (2000): A Diary Study of Information Capture in Working Life. In: Turner, Thea, Szwillus, Gerd, Czerwinski, Mary, Peterno, Fabio and Pemberton, Steven (eds.) Proceedings of the ACM CHI 2000 Human Factors in Computing Systems Conference April 1-6, 2000, The Hague, The Netherlands. pp. 438-445*

*Buxton, Bill (2001): Less is More (More or Less): Uncommon Sense and the Design of Computers. In: Denning, Peter J. (ed.). "The Invisible Future: The Seamless Integration of Technology Into Everyday Life". McGraw-Hill Companies*

*Bødker, Susanne (2006): When second wave HCI meets third wave challenges. In: Proceedings of the Fourth Nordic Conference on Human-Computer Interaction 2006. pp. 1-8*

*Bødker, Susanne (2006): When second wave HCI meets third wave challenges. In: Mørch, Anders, Morgan, Konrad, Bratteteig, Tone, Ghosh, Gautam and Svanaes, Dag (eds.) NordiCHI Nordic Conference on Human-Computer Interaction October 14-18, 2006, Oslo, Norway. pp. 1-8*

*Bødker, Susanne and Klokmose, Clemens N. (2011): The Human-Artifact Model: An Activity Theoretical Approach to Artifact Ecologies. In Human–Computer Interaction, 26 (4) pp. 315-371*

*Chalmers, Matthew (2004): A Historical View of Context. In Computer Supported Cooperative Work, 13 (3) pp. 223-247*

*Chen, Brian X. (2010). What the iPad means for the future of computing. Retrieved 8 November 2012 from wired.com: http://www.wired.com/gadgetlab/2010/02/ipad-future/*

*Chen, Guanling and Kotz, David (2000). A survey of context-aware mobile computing research. Dartmouth College Hanover http://dl.acm.org/citation.cfm?id=867843*

*Cheverst, Keith, Davies, Nigel, Mitchell, Keith, Friday, Adrian and Efstratiou, Christos (2000): Developing a Context-Aware Electronic Tourist Guide: Some Issues and Experiences. In: Turner, Thea, Szwillus, Gerd, Czerwinski, Mary, Peterno, Fabio and Pemberton, Steven (eds.) Proceedings of the ACM CHI 2000 Human Factors in Computing Systems Conference April 1-6, 2000, The Hague, The Netherlands. pp. 17-24*

*Cheverst, Keith, Davies, Nigel, Mitchell, Keith and Efstratiou, Christos (2001): Using Context as a Crystal Ball: Rewards and Pitfalls. In Personal and Ubiquitous Computing, 5 (1) pp. 8-11*

*Crabtree, Isac B. and Rhodes, Bradley J. (1998): Wearable Computing and the Remembrance Agent. In BT Technology Journal, 16 (3) pp. 118-124*

*de Sá, Marco and Carrico, Luis (2009): A mobile tool for in-situ prototyping. In: Proceedings of 11th Conference on Human-computer interaction with mobile devices and services 2009. p. 20*

*Dey, Anind K. (2001): Understanding and Using Context. In Personal and Ubiquitous Computing, 5 (1) pp. 4-7*

*Dix, Alan J., Rodden, Tom, Davies, Nigel, Trevor, Jonathan, Friday, Adrian and Palfreyman, Kevin (2000): Exploiting Space and Location as a Design Framework for Interactive Mobile Systems. In ACM Transactions on Computer-Human Interaction, 7 (3) pp. 285-321*

*Dix, Alan J., Finlay, Janet E., Abowd, Gregory D. and Beale, Russell (2004): Human-Computer Interaction (3rd Edition). Prentice Hall*

*Dourish, Paul (2004): What we talk about when we talk about context. In Personal and Ubiquitous Computing, 8 (1) pp. 19-30*

*Dourish, Paul (2001): Seeking a Foundation for Context-Aware Computing. In Human-Computer Interaction, 16 (2) pp. 229-241*

*Edwards, W. Keith (2005): Putting computing in context: An infrastructure to support extensible context-enhanced collaborative applications. In ACM Trans. Comput.-Hum. Interact., 12 (4) pp. 446-474*

*Feiner, Steven K., MacIntyre, Blair and Seligmann, Doree Duncan (1992): Annotating the real world with knowledge--based graphics on a see--through head--mounted display. In: Graphics Interface 92 May 11-15, 1992, Vancouver, British Columbia, Canada. pp. 78-85*

*Fling, Brian (2009): Mobile Design and Development: Practical concepts and techniques for creating mobile sites and web apps (Animal Guide). OReilly Media*

*Fortunati, Leopoldina (2001): The Mobile Phone: An Identity on the Move. In Personal and Ubiquitous Computing, 5 (2) pp. 85-98*

*Frederick, Gail and Lal, Rajesh (2010): Beginning Smartphone Web Development: Building JavaScript, CSS, HTML and Ajax-based Applications for iPhone, Android, Palm Pre, BlackBerry, Windows Mobile and Nokia S60. Apress*

*Green, Nicola, Harper, Richard H. R., Murtagh, Gerald and Cooper, Geoff (2001): Configuring the Mobile User: Sociological and Industry Views. In Personal and Ubiquitous Computing, 5 (2) pp. 146-156*

*Greenfield, Adam (2006): Everyware: The Dawning Age of Ubiquitous Computing. New Riders Publishing*

*Gye, Lisa (2007): Picture this: the impact of mobile camera phones on personal photographic practices. In Journal of Media and Cultural Studies, 21 (2) pp. 279-288*

*Hagen, Penny, Robertson, Toni, Kan, Melanie and Sadler, Kirsten (2005): Emerging research methods for understanding mobile technology use. In: Proceedings of OZCHI05, the CHISIG Annual Conference on Human-Computer Interaction 2005. pp. 1-10*

*Harrison, Steve, Art, Courtesy, Tatar, Deborah and Sengers, Phoebe (2007): The Three Paradigms of HCI. In: Begole, Bo, Payne, Stephen, Churchill, Elizabeth, Amant, Rob St., Gilmore, David and Rosson, Mary B. (eds.) Proceedings of Computer/Human Interaction 2007 April 28 – May 3, 2007, San Jose, USA. pp. 1-18*

*Helal, Abdelsalam A., Haskell, Bert, Carter, Jeffery L., Brice, Richard, Woelk, Darrell and Rusinkiewicz, Marek (1999): Any Time, Anywhere Computing: Mobile Computing Concepts and Technology (The Springer International Series in Engineering and Computer Science). Springer*

*Hinckley, Ken and Horvitz, Eric (2001): Toward more sensitive mobile phones. In: Marks, Joe and Mynatt, Elizabeth D. (eds.) Proceedings of the 14th annual ACM symposium on User interface software and technology November 11 - 14, 2001, Orlando, Florida. pp. 191-192*

*Hinckley, Ken, Pierce, Jeffrey S., Horvitz, Eric and Sinclair, Mike (2005): Foreground and background interaction with sensor-enhanced mobile devices. In ACM Trans. Comput.-Hum. Interact., 12 (1) pp. 31-52*

*Hinckley, Ken, Pierce, Jeff, Sinclair, Mike and Horvitz, Eric (2000): Sensing Techniques for Mobile Interaction. In: Ackerman, Mark S. and Edwards, Keith (eds.) Proceedings of the 13th annual ACM symposium on User interface software and technology November 06 - 08, 2000, San Diego, California, United States. pp. 91-100*

*Hosbond, Jens Henrik (2005): Mobile Systems Development Challenges, Implications and Issues. In: Krogstie, John, Kautz, Karlheinz and Allen, David (eds.) IFIP International Working Conference on Mobile Information Systems December 6-7, 2005, Leeds, United Kingdom . pp. 279-286*

*Høegh, Rune T., Kjeldskov, Jesper, Skov, Mikael B. and Stage, Jan (2008): A Field Laboratory for Evaluating In Situ. In: Klinger, Kristin, Roth, Kristin, Neidig, Jennifer, Reed, Sara, Langel, Joy, Smalley, Katie and Thor, Angela (eds.). "Handbook of Research on User Interface Design and Evaluation for Mobile Technology". IGI Globalpp. 982-996*

*Jameson, Anthony (2001): Modeling both the Context and the User. In Personal and Ubiquitous Computing, 5 (1) pp. 29-33*

*Johnson, Peter (1998). Usability and Mobility; Interactions on the move. Department of Computing Science http://www.dcs.gla.ac.uk/~johnson/papers/mobile/HCIMD1.html*

*Jones, Matt and Marsden, Gary (2006): Mobile interaction design. John Wiley and Sons*

*Jones, Quentin, Grandhi, Sukeshini A., Terveen, Loren and Whittaker, Steve (2004): People-to-People-to-Geographical-Places: The P3 Framework for Location-Based Community Systems. In Computer Supported Cooperative Work, 13 (3) pp. 249-282*

*Jung, Heekyoung, Stolterman, Erik A., Ryan, Will, Thompson, Tonya and Siegel, Marty (2008): Toward a framework for ecologies of artifacts: how are digital artifacts interconnected within a personal life?. In: Proceedings of the Fifth Nordic Conference on Human-Computer Interaction 2008. pp. 201-210*

*Kaikkonen, Anne, Kekäläinen, Aki, Cankar, Mihael, Kallio, Titti and Kankainen, Anu (2005): Usability Testing of Mobile Applications: A Comparison between Laboratory and Field Testing. In Journal of Usability Studies, 1 (1) pp. 4-17*

*Karapanos, Evangelos, Zimmerman, John, Forlizzi, Jodi and Martens, Jean-Bernard (2009): User experience over time: an initial framework. In: Proceedings of ACM CHI 2009 Conference on Human Factors in Computing Systems 2009. pp. 729-738*

*Kay, Alan C. (1972): A Personal Computer for Children of All Ages. In Proceedings of the ACM national conference,*

*Kindberg, Tim, Spasojevic, Mirjana, Fleck, Rowanne and Sellen, Abigail (2005): The ubiquitous camera: an in-depth study of camera phone use. In IEEE Pervasive Computing, 4 (2) pp. 42-50*

*Kjeldskov, Jesper and Howard, Steve (2004): Envisioning Mobile Information Services: Combining User- and Technology-Centered Design. In: Masoodian, Masood, Jones, Steve and Rogers, Bill (eds.) Computer Human Interaction 6th Asia Pacific Conference - APCHI 2004 June 29 - July 2, 2004, Rotorua, New Zealand. pp. 180-190*

*Kjeldskov, Jesper and Paay, Jeni (2005): Just-for-us: a context-aware mobile information system facilitating sociality. In: Proceedings of 7th conference on Human-computer interaction with mobile devices and services 2005. pp. 23-30*

*Kjeldskov, Jesper and Paay, Jeni (2010): Indexicality: Understanding mobile human-computer interaction in context. In ACM Transactions on Computer-Human Interaction, 17 (4) p. 14*

*Kjeldskov, Jesper and Paay, Jeni (2006): Public Pervasive Computing: Making the Invisible Visible. In IEEE Computer, 39 (9) pp. 60-65*

*Kjeldskov, Jesper and Skov, Mikael B. (2007b): Studying Usability In Sitro: Simulating Real World Phenomena in Controlled Environments. In International Journal of Human-Computer Interaction, 22 (1) pp. 7-36*

*Kjeldskov, Jesper and Skov, Mikael B. (2007a): Exploring context-awareness for ubiquitous computing in the healthcare domain. In Personal and Ubiquitous Computing, 11 (7) pp. 549-562*

*Kjeldskov, Jesper and Stage, Jan (2012): Combining ethnography and object-orientation for mobile interaction design: Contextual richness and abstract models. In International Journal of Human-Computer Studies, 70 (3) pp. 197-217*

*Kjeldskov, Jesper and Stage, Jan (2006): Exploring 'Canned Communication' for coordinating distributed mobile work activities. In Interacting with Computers, 18 (6) pp. 1310-1335*

*Kjeldskov, Jesper and Stage, Jan (2004): New techniques for usability evaluation of mobile systems. In International Journal of Human-Computer Studies, 60 (5) pp. 599-620*

*Kjeldskov, Jesper, Skov, Mikael B., Als, Benedikte S. and Høegh, Rune Thaarup (2004): Is It Worth the Hassle? Exploring the Added Value of Evaluating the Usability of Context-Aware Mobile Systems in the Field. In: Brewster, Stephen A. and Dunlop, Mark D. (eds.) Mobile Human-Computer Interaction - Mobile HCI 2004 - 6th International Symposium September 13-16, 2004, Glasgow, UK. pp. 61-73*

*Kjeldskov, Jesper, Skov, Mikael B., Paay, Jeni and Pathmanathan, Rahuvaran (2012): Using Mobile Phones to Support Sustainability: A Field Study of Residential Electricity Consumption. In: Kjeldskov, Jesper and Skov, Mikael B. (eds.) Proceedings of the ACM CHI 2012 Conference on Human Factors in Computing Systems 2012, Austin, Texas. pp. 2347-2356*

*Kjeldskov, Jesper, Gibbs, Martin R., Vetere, Franks, Howard, Steve, Pedell, Sonja, Mecoles, Karen and Bunyan, Marcus (2004b): Using Cultural Probes to Explore Mediated Intimacy. In Australasian Journal of Information Systems, 11 (2) pp. 102-115*

*Kjeldskov, Jesper, Graham, Connor, Pedell, Sonja, Vetere, Frank, Howard, Steve, Balbo, S. and Davies, J. (2005): Evaluating the usability of a mobile guide: the influence of location, participants and resources. In Behaviour & IT, 24 (1) pp. 51-65*

*Kjeldskov, Jesper, Christensen, Claus M. and Rasmussen, Klaus K. (2010): GeoHealth: a location-based service for home healthcare workers. In J. Location Based Services, 4 (1) pp. 3-27*

*Kostakos, Vassilis, Nicolai, Tom, Yoneki, Eiko, O'Neill, Eamonn, Kenn, Holger and Crowcroft, Jon (2009): Understanding and measuring the urban pervasive infrastructure. In Personal and Ubiquitous Computing, 13 (5) pp. 355-364*

*Lauesen, Soren (2004): User Interface Design: A Software Engineering Perspective. Addison-Wesley*

*Laurel, Brenda K. (ed.) (1990): The Art of Human-Computer Interface Design. Reading, MA, Addison-Wesley Publishing*

*Lindholm, Christian and Keinonen, Turkka (2003): Mobile Usability: How Nokia Changed the Face of the Mobile Phone. McGraw-Hill Professional*

*Lindley, Sian E., Harper, Richard, Randall, Dave, Glancy, Maxine and Smyth, Nicola (2009): Fixed in time and "time in motion": mobility of vision through a SenseCam lens. In: Proceedings of 11th Conference on Human-computer interaction with mobile devices and services 2009. p. 2*

*Ling, Rich (2001): "We Release Them Little by Little": Maturation and Gender Identity as Seen in the Use of Mobile Telephony. In Personal and Ubiquitous Computing, 5 (2) pp. 123-136*

*Little, Linda and Briggs, Pam (2009): Private whispers/public eyes: Is receiving highly personal information in a public place stressful?. In Interacting with Computers, 21 (4) pp. 316-322*

*Luff, Paul and Heath, Christian (1998): Mobility in Collaboration. In: Poltrock, Steven and Grudin, Jonathan (eds.) Proceedings of the 1998 ACM conference on Computer supported cooperative work November 14 - 18, 1998, Seattle, Washington, United States. pp. 305-314*

*McCullough, Malcolm (2005): Digital Ground: Architecture, Pervasive Computing, and Environmental Knowing. The MIT Press*

McCullough, Malcolm (2004): Digital Ground: Architecture, Pervasive Computing, and Environmental Knowing. The MIT Press

Mikkonen, M., Väyrynen, S., Ikonen, V. and Heikkilä, M. O. (2002): User and Concept Studies as Tools in Developing Mobile Communication Services for the Elderly. In Personal and Ubiquitous Computing, 6 (2) pp. 113-124

Miller, Frederic P, Vendome, Agnes F. and McBrewster, John (2010): Digital Ecosystem. Beau-Bassin, Mauritius, Alphascript Publishing

Moggridge, Bill (2007): Designing Interactions. The MIT Press

Murphy, John, Kjeldskov, Jesper, Howard, Steve, Shanks, Graeme and Hartnell-Young, Elizabeth (2005): The converged appliance: "I love it... but I hate it". In: Proceedings of OZ-CHI05, the CHISIG Annual Conference on Human-Computer Interaction 2005. pp. 1-10

Nielsen, Jakob (2000). WAP Field Study Findings. Retrieved 8 November 2012 from useit.com: http://www.useit.com/alertbox/20001210.html

Nielsen, Christian Monrad, Overgaard, Michael, Pedersen, Michael Bach, Stage, Jan and Stenild, Sigge (2006): It's worth the hassle!: the added value of evaluating the usability of mobile systems in the field. In: Mørch, Anders I., Morgan, Konrad, Bratteteig, Tone, Ghosh, Gautam and Svanaes, Dag (eds.) Proceedings of the 4th Nordic Conference on Human-Computer Interaction 2006, Oslo, Norway, October 14-18, 2006 2006. pp. 272-280

Norman, Donald A. (1998): The Invisible Computer: Why Good Products Can Fail, the Personal Computer Is So Complex and Information Appliances Are the Solution. MIT Press

O'Hara, Kenton, Kjeldskov, Jesper and Paay, Jeni (2011): Blended interaction spaces for distributed team collaboration. In ACM Transactions on Computer-Human Interaction (TOCHI), 18 (1) pp. 3:1-3:28

Oulasvirta, Antti (2009): Field Experiments in HCI: Promises and Challenges. In: Saariluoma, Pertti and Isomäki, Hannakaisa (eds.). "Future Interaction Design II". London, United Kingdom: pp. 87-111

Oulasvirta, Antti and Nyyssönen, Tuomo (2009): Flexible Hardware Configurations for Studying Mobile Usability. In Journal of Usability Studies, 4 (2) pp. 93-105

Paay, Jeni (2008): From ethnography to interface design. In: Klinger, Kristin, Roth, Kristin, Neidig, Jennifer, Reed, Sara, Langel, Joy, Smalley, Katie and Thor, Angela (eds.). "Handbook of Research on User Interface Design and Evaluation for Mobile Technology". IGI Global

*Paay, Jeni and Kjeldskov, Jesper (2008b): Understanding the user experience of location-based services: five principles of perceptual organisation applied. In Journal of Location Based Services, 2 (4) pp. 267-286*

*Paay, Jeni and Kjeldskov, Jesper (2005): Understanding and modelling built environments for mobile guide interface design. In Behaviour and Information Technology, 24 (1) pp. 21-35*

*Paay, Jeni and Kjeldskov, Jesper (2008a): Situated Social Interactions: A Case Study of Public Places in the City. In Computer Supported Cooperative Work, 17 (2) pp. 275-290*

*Paay, Jeni, Sterling, Leon, Vetere, Frank, Howard, Steve and Boettcher, Anne (2009b): Engineering the social: The role of shared artifacts. In International Journal of Human-Computer Studies, 67 (5) pp. 437-454*

*Paay, Jeni, Kjeldskov, Jesper, Howard, Steve and Dave, Bharat (2009a): Out on the town: A socio-physical approach to the design of a context-aware urban guide. In ACM Transactions on Computer-Human Interaction, 16 (2) p. 7*

*Palen, Leysia, Salzman, Marilyn C. and Youngs, Ed (2000): Going Wireless: Behavior & Practice of New Mobile Phone Users. In: Kellogg, Wendy A. and Whittaker, Steve (eds.) Proceedings of the 2000 ACM conference on Computer supported cooperative work 2000, Philadelphia, Pennsylvania, United States. pp. 201-210*

*Perry, Mark, O'Hara, Kenton P., Sellen, Abigail, Brown, Barry and Harper, Richard (2001): Dealing with mobility: understanding access anytime, anywhere. In ACM Transactions on Computer-Human Interaction, 8 (4) pp. 323-347*

*Preece, Jennifer J., Rogers, Yvonne, Sharp, Helen, Benyon, David, Holland, Simon and Carey, Tom (1994): Human-Computer Interaction. Reading, Mass., Addison-Wesley Publishing*

*Preece, Jennifer J., Rogers, Yvonne and Sharp, Helen (2002): Interaction Design: Beyond Human-Computer Interaction. John Wiley and Sons*

*Ramsay, Marc and Nielsen, Jakob (2000). WAP Usability: Déjà Vu: 1994 All Over Again. Nielsen Norman Group http://www.nngroup.com/reports/wap/WAP_usability.pdf*

*Raskin, Jef (2000): The Humane Interface: New Directions for Designing Interactive Systems. Addison-Wesley Publishing*

*Reichl, Peter, Fröhlich, Peter, Baillie, Lynne, Schatz, Raimund and Dantcheva, Antitza (2007): The LiLiPUT prototype: a wearable lab environment for user tests of mobile telecommunication applications. In: Rosson, Mary Beth and Gilmore, David J. (eds.) Extended*

Abstracts Proceedings of the 2007 Conference on Human Factors in Computing Systems, CHI 2007, San Jose, California, USA, April 28 - May 3, 2007 2007. pp. 1833-1838

Rodden, Tom, Chervest, Keith, Davies, Nigel and Dix, Alan (1998). Exploiting Context in HCI Design for Mobile Systems. ACM http://www.dcs.gla.ac.uk/~johnson/papers/mobile/HCIMD1.html

Rogers, Yvonne, Sharp, Helen and Preece, Jenny (2011): Interaction Design: Beyond Human - Computer Interaction - third edition. Wiley

Rogers, Yvonne, Connelly, Kay, Tedesco, Lenore, Hazlewood, William R., Kurtz, Andrew, Hall, Robert E., Hursey, Josh and Toscos, Tammy (2007): Why It's Worth the Hassle: The Value of In-Situ Studies When Designing Ubicomp. In: Krumm, John, Abowd, Gregory D., Seneviratne, Aruna and Strang, Thomas (eds.) UbiComp 2007 Ubiquitous Computing - 9th International Conference September 16-19, 2007, Innsbruck, Austria. pp. 336-353

Rowland, Duncan, Flintham, Martin, Oppermann, Leif, Marshall, Joe, Chamberlain, Alan, Koleva, Boriana, Benford, Steve and Perez, Citlali (2009): Ubikequitous computing: designing interactive experiences for cyclists. In: Proceedings of 11th Conference on Human-computer interaction with mobile devices and services 2009. p. 21

Schilit, Bill N. and Theimer, Marvin M. (1994): Disseminating active map information to mobile hosts. In IEEE Network, 8 (5) pp. 22-32

Schmidt, Albrecht, Beigl, Michael and Gellersen, Hans-Werner (1999b): There is more to context than location. In Computers & Graphics, 23 (6) pp. 893-901

Schmidt, Albrecht, Aidoo, Kofi Asante, Takaluoma, Antti, Tuomela, Urpo, Laerhoven, Kristof van and Velde, Walter van de (1999): Advanced Interaction in Context. In: Gellersen, Hans-Werner (ed.) Handheld and Ubiquitous Computing - First International Symposium - HUC99 September 27-29, 1999, Karlsruhe, Germany. pp. 89-101

Sharp, Helen, Rogers, Yvonne and Preece, Jennifer J. (2007): Interaction Design: Beyond Human-Computer Interaction. John Wiley and Sons

Sharples, Mike, Corlett, Dan and Westmancott, Oliver (2002): The Design and Implementation of a Mobile Learning Resource. In Personal and Ubiquitous Computing, 6 (3) pp. 220-234

*Shneiderman, Ben (1997): Designing the User Interface: Strategies for Effective Human-Computer Interaction. (3rd ed.). Boston, MA, Addison-Wesley Publishing*

*Skov, Mikael B., Kjeldskov, Jesper, Paay, Jeni, Husted, Niels, Nørskov, Jacob and Pedersen, Kenneth (2012): Designing on-site: Facilitating Participatory Contextual Architecture with Mobile Phones. In Journal of Pervasive and Mobile Computing, 8 (6)*

*Vetere, Frank, Gibbs, Martin R., Kjeldskov, Jesper, Howard, Steve, Mueller, Florian, Pedell, Sonja, Mecoles, Karen and Bunyan, Marcus (2005): Mediating intimacy: designing technologies to support strong-tie relationships. In: Proceedings of ACM CHI 2005 Conference on Human Factors in Computing Systems 2005. pp. 471-480*

*Weilenmann, Alexandra (2001): Negotiating Use: Making Sense of Mobile Technology. In Personal and Ubiquitous Computing, 5 (2) pp. 137-145*

*Weiser, Mark (1991): The Computer for the 21st Century. In Scientific American, 265 (3) pp. 94-104*

*Weiss, Scott (2002): Handheld Usability. John Wiley and Sons*

*Wilde, Oscar (2010): The Decay Of Lying. Kessinger Publishing, LLC*

*Winograd, Terry (1996): Bringing Design to Software. ACM Press*

*Zuberec, Sarah (2000): The interaction design of Microsoft Windows CE. In: Bergman, Eric (ed.). "Information Appliances and Beyond: Interaction Design for Consumer Products (Interactive Technologies)". Morgan Kaufmann*

## About the author

### Jesper Kjeldskov

*© Jesper Kjeldskov*

Has also published under the name of:
"J. Kjeldskov"

Personal Homepage:
http://people.cs.aau.dk/~jesper/

Current place of employment:
Aalborg University, Centre for Socio-Interactive Design

Jesper Kjeldskov is Associate Professor in Human-Computer Interaction at Aalborg University's Department of Computer Science. Jesper's research interests are Interaction Design and User Experience with particular focus on mobile and ubiquitous technologies in non-work settings. Jesper has a cross-disciplinary background spanning the humanities, social sciences and computer science.

He has published more than 120 peer-reviewed papers within the area of Human-Computer Interaction including more than 20 articles in international journals. In 2008-09 Jesper was the research leader of a User Experience Research group in Sydney, Australia leading a multi-million dollar research project on Blended Interaction Spaces. Since 2009 he has revisited his interests in mobile HCI but also extended his work on domestic computing with new projects within sustainability, digital media, mediated relationships and interior architecture.

## Your notes and thoughts on chapter 9

Record your notes and thoughts on this chapter. If you want to share these thoughts with others online, go to the bottom of the page at:
http://www.interaction-design.org/encyclopedia/mobile_computing.html

**NOTES:**

# Chapter 10

# End-User Development

*by Margaret M. Burnett and Christopher Scaffidi.*

Computer users have rapidly increased in both number and diversity (Scaffidi et al 2005). They include managers, accountants, engineers, home makers, teachers, scientists, health care workers, insurance adjusters, salesmen, and administrative assistants. Many of these people work on tasks that rapidly vary on a yearly, monthly, or even daily basis. Consequently, their software needs are diverse, complex, and frequently changing. Professional software developers cannot directly meet all of these needs because of their limited domain knowledge and because their development processes are too slow.

End-user development (EUD) helps to solve this problem. EUD is "a set of methods, techniques and tools that allow users of software systems, who are acting as non-professional software developers, at some point to create, modify, or extend a software artifact" (Lieberman et al 2006). In particular, EUD enables end users to design or customize the user interface and functionality of software. This is valuable because end users know their own context and needs better than anybody else, and they often have real-time awareness of shifts in their respective domains. Through

EUD, end users can tune software to fit their requirements more closely than would be possible without EUD. Moreover, because end users outnumber professional software developers by a factor of 30-to-1 (Figure 10.1), EUD "scales out" software development activities by enabling a much larger pool of people to participate.

However, EUD is inherently different from traditional software development, and trying to support EUD by simply mimicking traditional approaches is often insufficient to produce successful results. End users usually do not have training in professionals' programming languages, formal development processes, or modeling and diagramming notations. Moreover, end users often lack the time or motivation to learn these traditional techniques, since end users usually write code in order to achieve a short- or medium-term goal rather than to create a durable software asset that will produce a continuing revenue stream. Consequently, supporting EUD requires providing appropriate tools, social structures, and development processes that are highly usable, quickly learned, and easily integrated into domain practice.

EUD overlaps with two similar concepts, end-user programming and end-user software engineering. End-user programming (EUP) enables end users to *create* their own programs (Ko et al 2011). This subset of EUD is the most mature from a research and practice perspective, so we focus a later section of this article on that portion of EUD. The difference between EUP and EUD is that EUD methods, techniques, and tools span the *entire* software development lifecycle, including modifying and extending software, not just the "create" phase.

The other related concept overlapping with EUD is end-user software engineering (EUSE). EUSE is a relatively new subset of EUD that began about a decade ago. Its emphasis is on the *quality* of the software end users create, modify, or extend; thus its research focuses on methods, techniques, and tools that promote the quality of such software. This area has arisen because of the ample evidence that the programs end users create are filled with expensive errors (Panko 1998; Burnett 2010; Ko et al 2011). We therefore focus on the EUSE subset of EUD in a later section of this article.

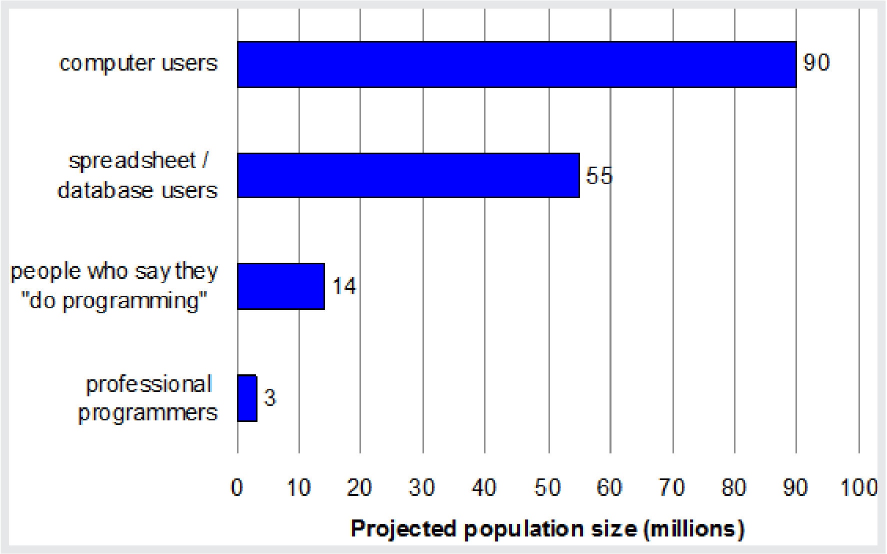



FIGURE 10.1: Projected population sizes for American workplaces in 2012, based on federal data (note that categories are not mutually exclusive).

*Courtesy of Christopher Scaffidi. Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.*

VIDEO 10.1: End-User-Development - Introduction to End-User Development.

*Courtesy of Rikke Friis Dam and Mads Soegaard. Copyright: CC-Att-ND (Creative Commons Attribution-NoDerivs 3.0 Unported). View full screen or download (480 KB)*

VIDEO 10.2 End-User-Development - Three Pieces of Advice Three Major Pitfalls and Three Best Practices.

*Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms. View full screen or download (0)*

VIDEO 10.3: End-User-Development - Business Value - Applicability in Industry.

*Courtesy of Rikke Friis Dam and Mads Soegaard. Copyright: CC-Att-ND (Creative Commons Attribution-NoDerivs 3.0 Unported). View full screen or download (480 KB)*


VIDEO 10.4: End-User-Development - Gender Differences in End-User Development.

*Courtesy of Rikke Friis Dam and Mads Soegaard. Copyright: CC-Att-ND (Creative Commons Attribution-NoDerivs 3.0 Unported). View full screen or download (960 KB)*

VIDEO 10.5: End-User-Development - Future Directions.

*Courtesy of Rikke Friis Dam and Mads Soegaard. Copyright: CC-Att-ND (Creative Commons Attribution-NoDerivs 3.0 Unported). View full screen or download (672 KB)*

## 10.1 THE BIRTH OF EUD

Prior to the 1980's, most computing occurred on mainframes controlled by professional developers in information systems departments. End users had little influence over the form and function of software running on a mainframe, which they generally viewed through simple terminal windows and controlled with simple textual commands (Figure 10.2). Information systems departments rarely had enough staff time to design and implement all of the software enhancements requested by users (Brancheau and Wetherbe 1987).

EUD grew out of a confluence of innovations embodied in the machines known as "microcomputers" (a term eventually replaced with "personal computer"). First, these machines were inexpensive enough that organizations could afford to provide each user with a machine. Having their own machines made it viable for users to modify ("tailor") the machine's software settings without impacting the computing environment of other users. Second, microcomputers had sufficient hardware power so that users could compile (or interpret) new code in languages such as Basic. This in turn provided infrastructure for end users to create new applications. Third, microcomputers soon came to include innovative new features such as the mouse and powerful graphics cards, which accelerated usability advances such as graphical user interfaces (GUIs) and direct manipulation; these advances, in turn, opened up the possibility of novel programming tools specifically designed to meet the needs of users.

Spreadsheets were the first major EUD programming environment made possible by these innovations (Bricklin et al 1979), beginning with VisiCalc (Figure 10.3), then continuing with Lotus 1-2-3 and Excel. Although users of spreadsheet systems may not think of themselves as "doing programming," spreadsheet systems are programming environments because their formulas are first-order functional programs (Jones et al 2003). In such programs, the formulas can refer to input "variables" (cell names) and the results of the formulas are computed output values. The availability of spreadsheet software was a major factor in spurring early demand for microcomputers (Ichbiah 1993).

Newer technologies such as the web and mobile computing have since opened up increasingly diverse and powerful opportunities for end users to create and tailor software.


Figure 10.2: Terminal interface presenting a fixed menu, where users type a number to indicate a menu selection.

FIGURE 10.3: Visicalc circa 1980.



## 10.2 TAILORING

Tailoring is any “activity to modify a computer application within its context of use” (Won et al 2006). Tailoring can be a simple or complex activity. At each increased level of complexity, users have more ability to redesign the interaction and functionality of an application. At the most basic level, tailoring encompasses specifying parameters to an existing application in a way that changes its behavior at a high level of granularity. For example, a person might use a graphical user interface to indicate which features of a spreadsheet editor should be visible (Figure 10.4) or how a word processor should respond to various inputs (Figure 10.5). Once tailoring begins to involve creating full-fledged programs in order to extend the functionality of an application, the activity seamlessly encompasses end-user programming (below). For instance, a user might create scripts called “macros” that manipulate the buttons, text, or other graphical user interface elements within an application. Such macros can extend applications with new functionality or make existing functionality more usable (Figure 10.6). Researchers have proposed a variety of component-based frameworks that can be used to implement easily-tailored applications (Won et al 2006). For instance, the "Selection" object referenced in Figure 10.6 is actually a component representing the region of text that is currently highlighted by the user in the word processor. The component-based framework makes it possible for an interpreter to manipulate the highlighted text in response to macros' instructions.

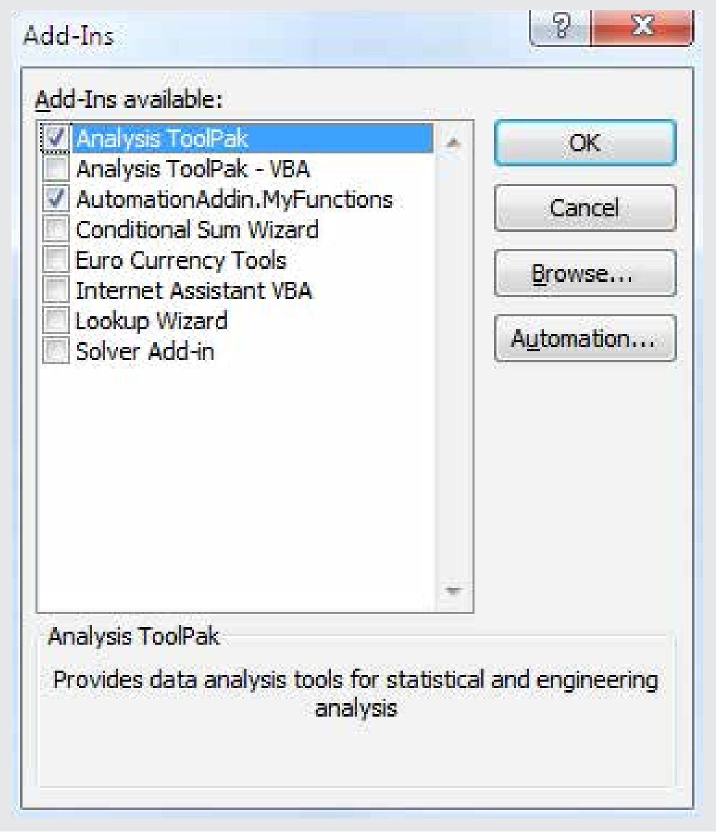



Figure 10.4: Screen in Microsoft Excel for tailoring which features should be activate and visible.

*Courtesy of Christopher Scaffidi. Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.*

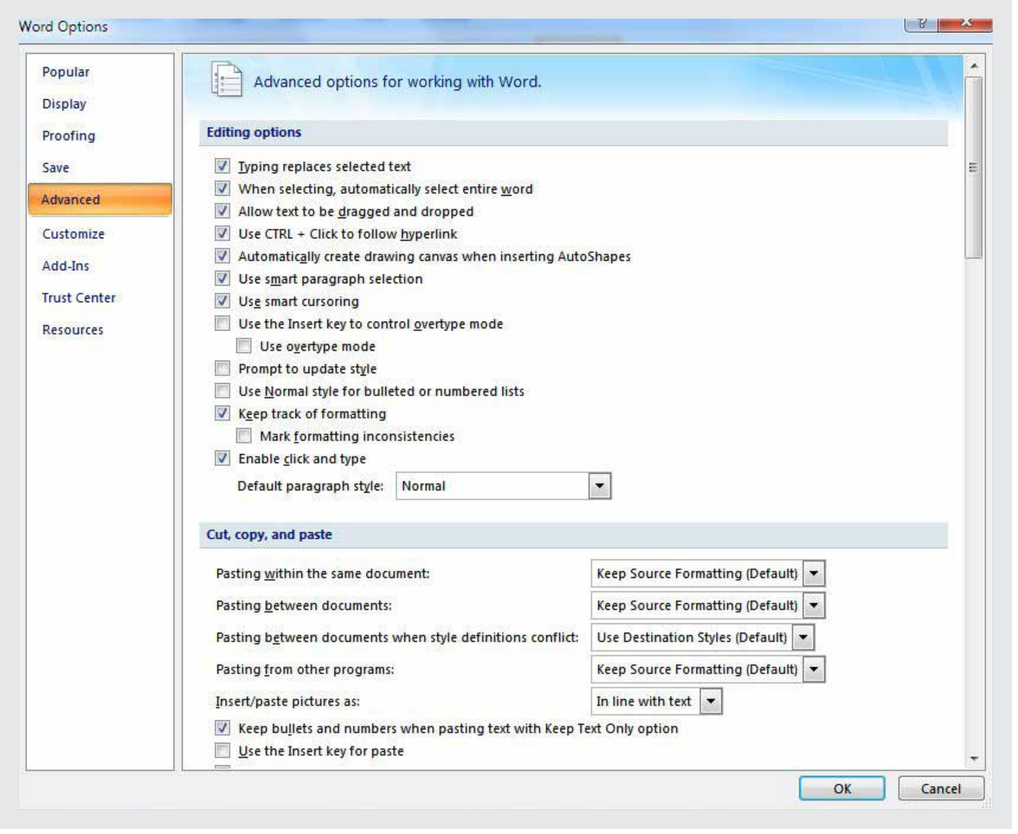



FIGURE 10.5: Screen in Microsoft Word for tailoring how the application handles clicks on hyperlinks, copy/paste commands, and other kinds of user input.

*Courtesy of Christopher Scaffidi. Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.*

```
Sub PasteSpecial()
Selection.PasteSpecial DataType:=wdPasteText
End Sub
```

```
Sub SmallCaps()
'
' SmallCaps Macro
'
'
    With Selection.Font

        .SmallCaps = True

    End With
End Sub
```

```
Sub Subscript()
'
' Subscript Macro
'
'
    With Selection.Font
        If (.Superscript) Then
        .Superscript = False
        .Subscript = False
        ElseIf (.Subscript) Then
        .Subscript = False
        .Superscript = True
        Else
        .Subscript = True
        .Superscript = False
        End If


    End With
End Sub
```

Figure 10.6: Using a code editor to create three Microsoft Word text-manipulation macros that are linked (through a separate screen) to keyboard shortcuts; for example, the first macro has a single instruction indicating that the application should convert whatever is on the system clipboard into a textual representation and then paste it into the document.

## 10.3 END-USER PROGRAMMING (EUP)

End-user programming (EUP) is defined as “programming to achieve the result of a program, rather than the program itself” (Ko et al 2011). In EUP, the developer’s goal is to actually *use* the program; this contrasts with professional programming, where the goal is to create a program for *other* people to use, often in exchange for monetary compensation. The programs created through EUP can be extensions of existing applications (as in Figure 10.6, above), or they can be new applications that run separately from existing applications. End users can perform EUP through a wide range of interaction styles (Nardi 1993), as we discuss next.

### 10.3.1 Programming using visual attributes

In environments supporting the visual programming style of interaction, at least some of a program’s semantics is expressed through the visual layout of the program. For example, the grid-like arrangement of cells in a spreadsheet carries a certain semantics; specifically, cells that are vertically or horizontally aligned with one another are part of a composite object defined solely based on the visual layout of cells (e.g, the range B:B references all of the second column in Microsoft Excel). In a visual language, semantics can hypothetically be encoded in many attributes of a visual representation, such as position, color, size, and intersection with other shapes. As another example, Figure 10.7 shows a visual language where each instruction is a colored block whose color indicates what kind of instruction it is and whose shape indicates what other blocks can appear next or before this block. Figure 10.8 shows a third example of a visual language. As with many visual languages, the LabView programming tool allows users to drag and drop these shapes using the GUI. Another common way of interacting through a visual language is with a form (Figure 10.9). In such an interface, the user cannot freely drag and drop shapes but rather must make selections from pre-defined fields.

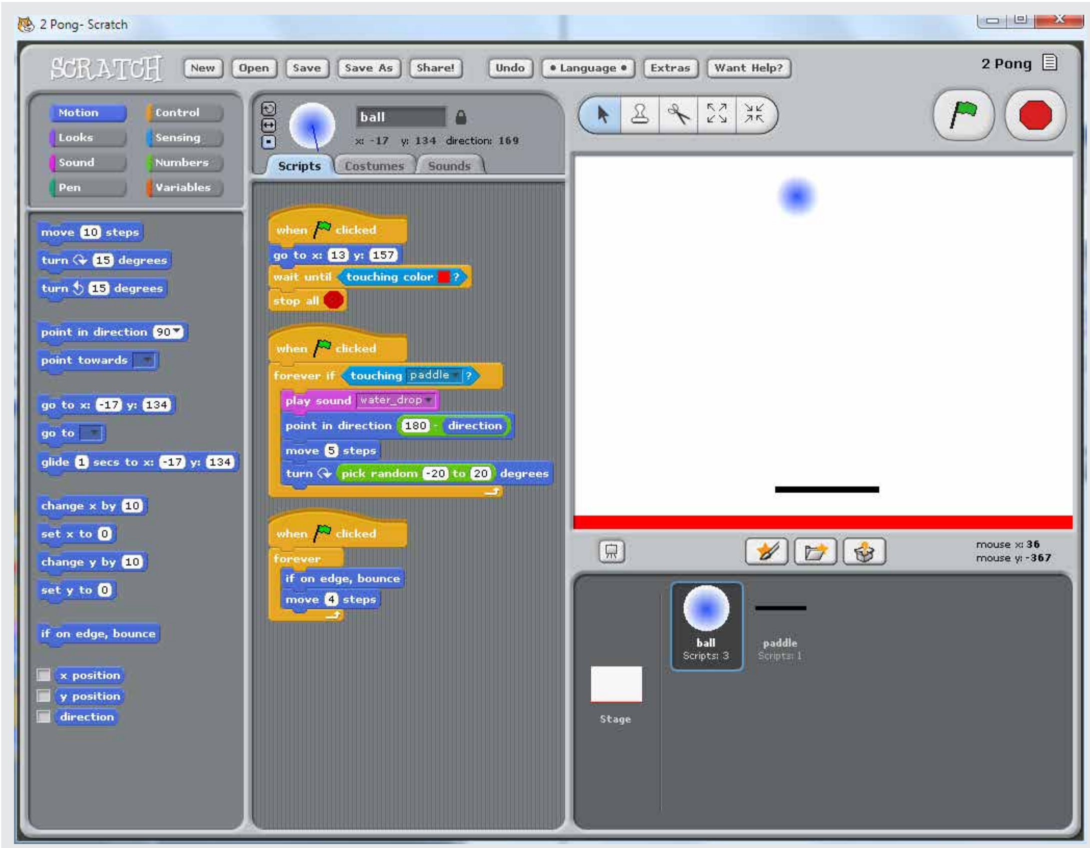



Figure 10.7: Visual interface for editing the three scripts for the ball in a 'pong' game animation. The programmer lays out sprites on the right; clicking a sprite brings up its scripts for editing in the center. Primitives can be dragged-and-dropped from the toolbox at left. (Resnick et al 2009).

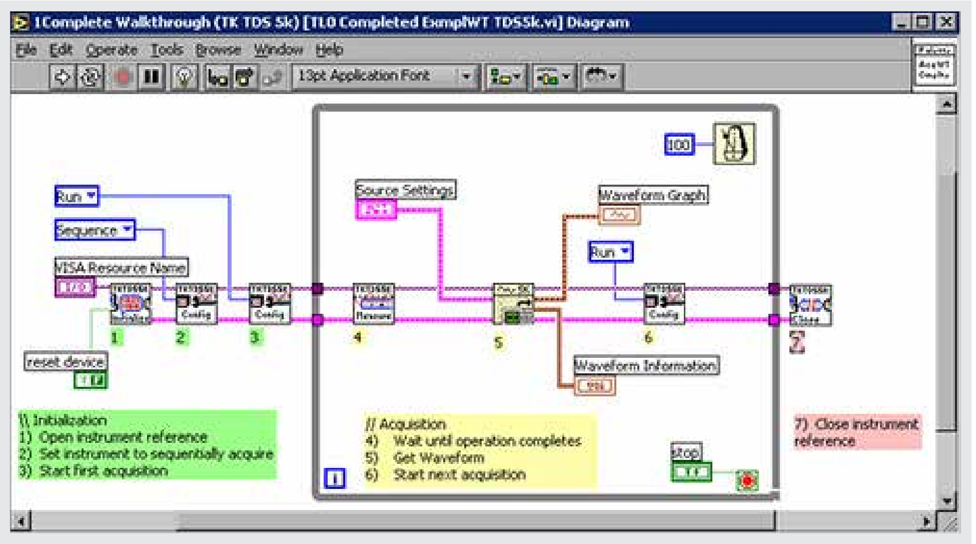



Figure 10.8: LabVIEW programming language for creating circuit simulations and other programs. Each box represents a computational component, while lines indicate flow of data (similar to wires carrying signals).

*Courtesy of Sam Shearman. Copyright: CC-Att-SA-2 (Creative Commons Attribution-ShareAlike 2.0 Unported).*

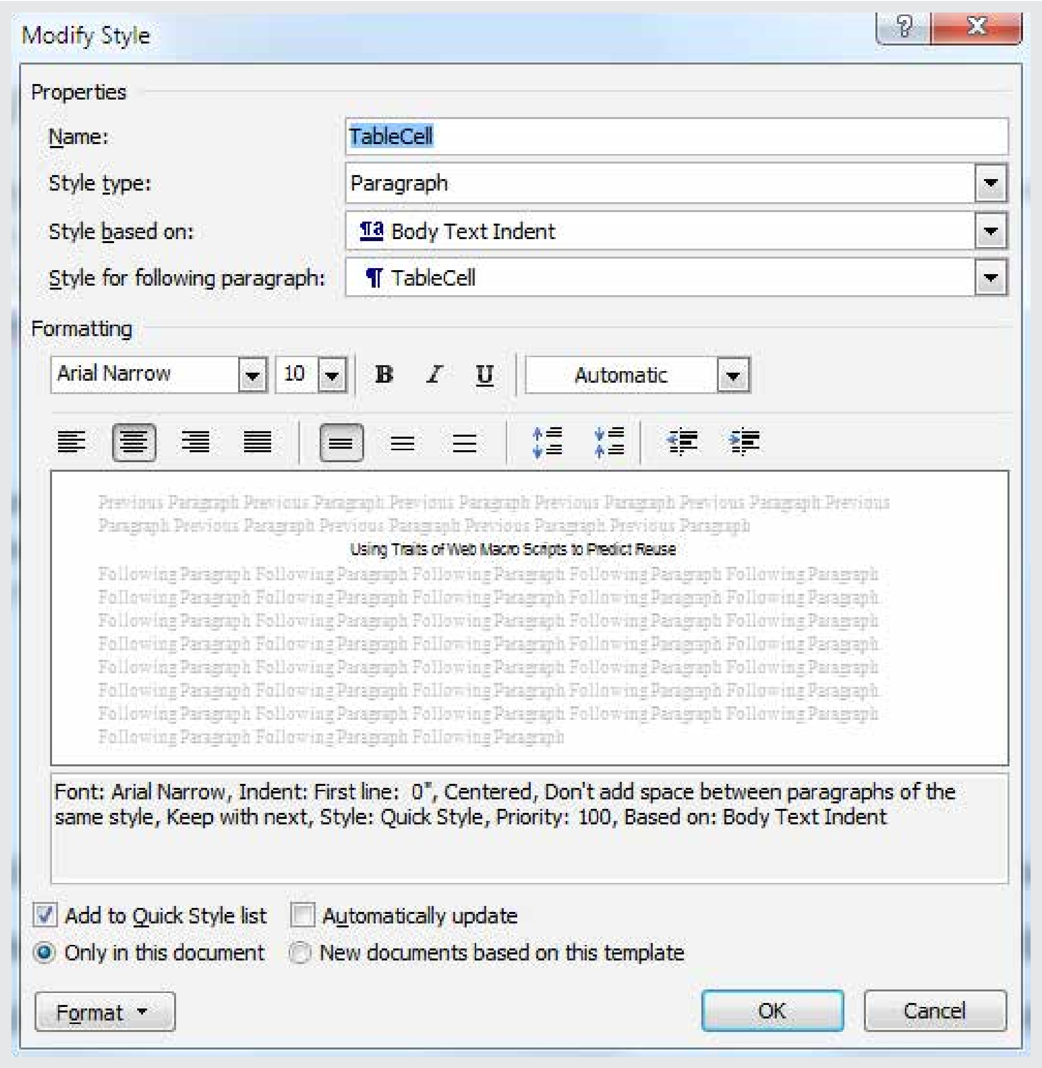



FIGURE 10.9: User interface in Microsoft Word for creating a style, which is a set of formatting instructions that will be applied to multiple labeled regions in the document.

### 10.3.2 Programming-by-demonstration (PBD)

Programming-by-demonstration (PBD), sometimes called programming-by-example, is a programming technique whereby the user demonstrates the new program's logic, from which the programming environment infers a program representing that logic. Some PBD systems are able to deductively infer the entire program, while others deduce what they can and ask the user for help for the rest (Cypher 1993). PBD-based tools are available for creating animations (Repenning and Perrone 2000; McDaniel and Myers 1999) and many other kinds of programs (Cypher 1993). One problem with PBD has been representing the final program in a form useful to the user (Cypher 1993; Yang et al 1997), to enable the end-user developer to review, test, and debug the program. Thus, PBD is often used in combination with visual or textual languages.

For example, a user could create a Microsoft Word macro (like those shown in Figure 10.6) through PBD. The user would first click a button or menu item indicating that the application should start watching the user's actions. The user would then use the GUI to demonstrate the desired behavior for the macro; for example, the user might use a series of menu items and dialog windows to paste the system clipboard as text. The user would click the "stop recording" button so that Microsoft Word stops watching the user's actions. At that point, the application would generate a macro containing VBScript instructions for repeating the demonstrated actions. The user could give the macro a keyboard shortcut and a name, if desired, in order to simplify running or editing it later. In addition, the user might want to edit the macro's instructions so that they perform a task slightly different than the one that was demonstrated, particularly if the user wants the macro to complete a task that is impossible with the existing GUI (rather than merely hard-to-use). In this way, the user could add completely new functionality to an application, with PBD serving to provide a starting point for another approach to programming.

### 10.3.3 Programming-by-specification

Programming-by-specification is an interaction style where the user describes a desired program, and a tool then generates the program for the user. As in PBD, the generated program can then be represented to facilitate review and customization by the user. For example, Liu and Lieberman (2005) implemented a system that accepts a specification in natural language and generates a corresponding program written in Python. A key limitation of this approach, as with inference-based PBD approaches, is that it is difficult for a user to predict what program will be generated from any particular input. Another is that, as with PBD, representation can be a difficult issue. For example, if the input interaction is done with English but the output is a traditional programming language such as Python, the end user must be fluent in both languages. Another limitation is that the programming tool can often correctly process only a limited range of inputs. This restricts the usefulness of the tool and also makes it difficult for a user to predict whether (and how) some particular input will be "understood" by the tool (e.g, could the Liu and Lieberman tool mentioned above generate games, and if so, what kinds of games?) In order to make the bounds of a tool's input language more obvious to users, some systems provide a forms-based visual interface (Figure 10.10) rather than a textual interface, thereby restricting users' specifications to only those that can actually be handled by the tool (Scaffidi et al 2009).

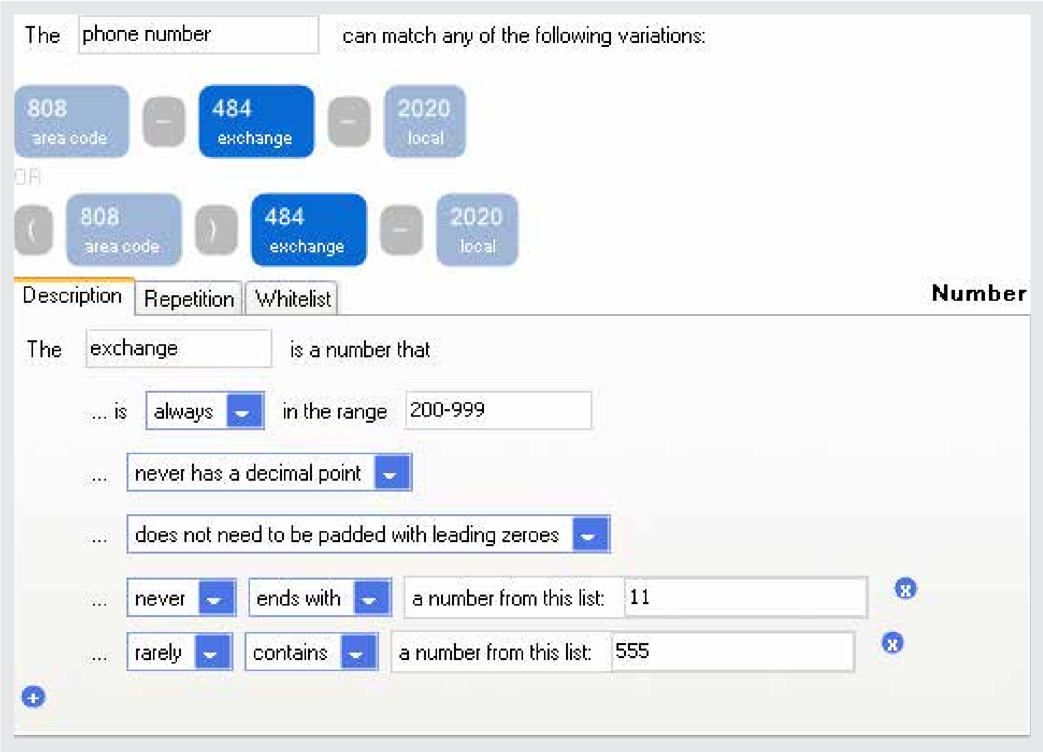



Figure 10.10: Visual specification of what a phone number looks like; from this specification, the tool generates code that can check whether a particular string matches the specification (Scaffidi 2009).

*Courtesy of Christopher Scaffidi. Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.*

## 10.3.4 Programming with text

Programming with text is the most traditional interaction technique for programming, and for a time, some believed that this style of programming would not be appropriate for EUP. However, as the previous examples have shown, most programming environments that support other interaction styles also include text to some extent. As another example, Figure 10.11 shows the textual language that the CoScripter tool uses to represent a web macro, which is a script that directs a web browser to navigate the web and manipulate web sites in a particular way (Leshed et al 2008); such a web macro is typically created through PBD and then customized in textual form as needed. Despite the proliferation of alternative interaction styles, text remains widely used because of its conciseness and effectiveness for communicating abstract concepts.

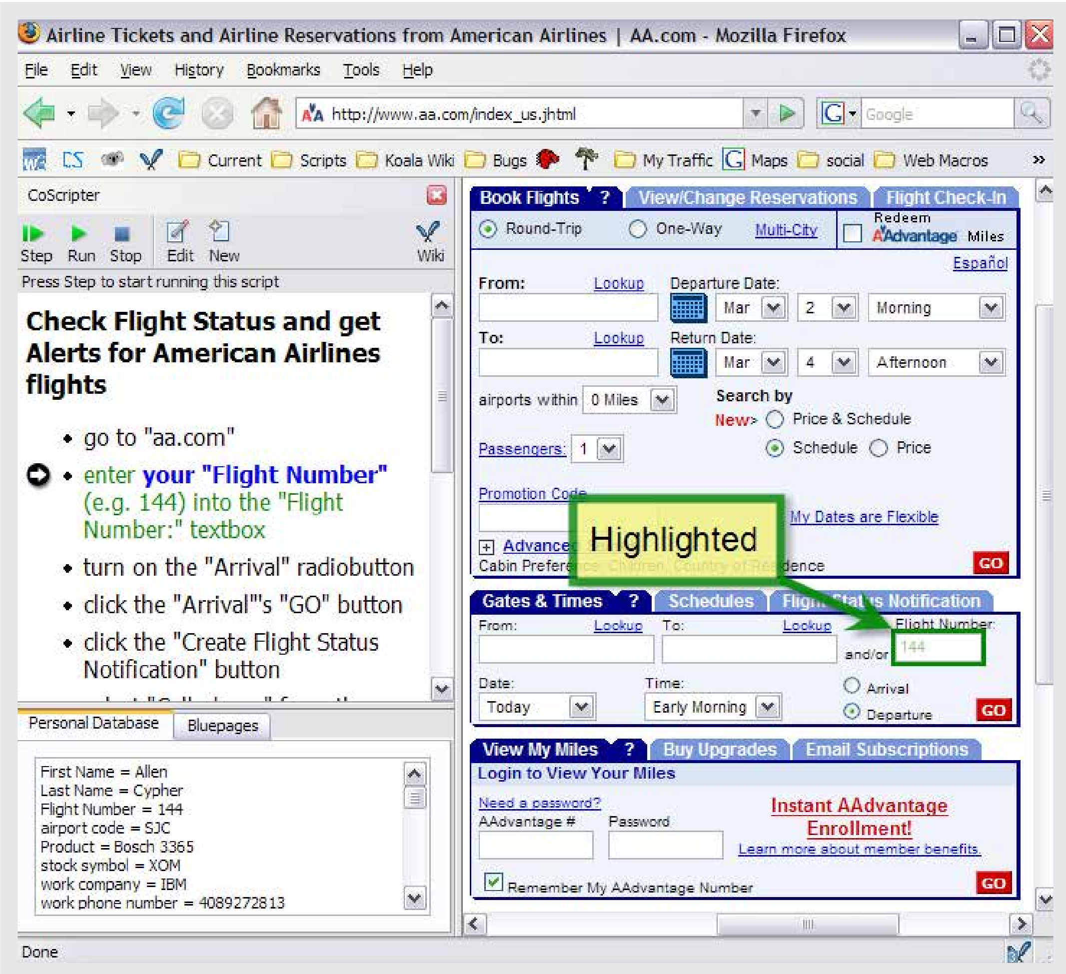



Figure 10.11: Using the CoScripter programming tool to edit a web macro that tells the browser to look up information on the American Airlines website. The macro's execution has been paused at the second instruction (left), which instructs CoScripter to highlight Flight Number on the web page (right) and fill it in from the user's 'Personal Database' configuration file (lower left). Scaffidi et al 2010.

## 10.4 END-USER SOFTWARE ENGINEERING (EUSE)

End-user software engineering (EUSE) is defined as "end-user programming involving systematic and disciplined activities that address software quality issues" (Ko et al 2011). Attention to quality is important for EUP because poorly-written software can cause data loss, security breaches, financial loss, or even physical harm, even when the software is created by end-user developers. The software qualities relevant to EUSE are the same as those of interest to professional developers who sell their products. These qualities include reliability, performance, maintainability, reusability, privacy, and security. Some qualities, such as maintainability and reusability, only become apparent after a program has been written and in operation for some time. Thus, EUSE combines the goal of EUP, which focuses on enabling end users to create software, with the concern for *quality* of that software across its *entire lifecycle*. This lifecycle includes requirements, design, verification, debugging and code reuse (in addition to actual implementation, which has already been described above in the context of EUP tools).

### 10.4.1 Requirements and design

Requirements describe what a program should do, and design refers to determining how a program should do it. For example, a requirement might be that a program should be able to sort a list of mailing addresses, and its design might detail the sorting algorithm to be used.

Examples of requirements (goals) in EUD include personalizing the way that an application or computer behaves, automating time-consuming tasks, performing computations that are hard to do accurately by hand, or communicating information (Ko et al 2011; Blackwell 2004; Blackwell 2006; Rosson et al 2002). Getting these requirements right is a critical aspect from the perspective of EUSE, because of its emphasis on quality. Professional developers are expected to investigate, document, and refine requirements before they start to design or code an application. For example, they might try to identify inconsistencies in requirements by applying one of several painstaking techniques (e.g, stakeholder review or formal modeling). In contrast, end users often live in their domain every day and know it very well, so they often already have an idea about what the requirements, and do not do any extra work to arrive at them, document them, or check them.

However, sometimes end users do not know the requirements in advance and do not aspire to a "design" per se; they may expect these matters to be clarified as the program's implementation evolves (Costabile et al 2006; Fischer and Giaccardi 2006; Morch and Mehandjiev 2000; Segal 2007). (Professional developers sometimes do not know the requirements in advance either, but they are expected to take steps to deal with that situation, such as employing an iterative method that fills in requirements as prototypes evolve, rather than entirely omitting the concept and moving on.) In this case, end-user developers may jump directly into coding without taking the time to document their requirements or look for inconsistencies (Rosson et al 2010). A related situation is that, because of the tight coupling of EUD to a domain, external shifts in the domain can cause evolution in requirements; for example, changes to accounting rules might require a financial analyst to compute different data, which might in turn cause modifications to an existing spreadsheet. Due to its highly iterative nature, EUD requirements-refinement has been likened to a form of highly agile programming (Lieberman et al 2006).

Thus, end-user programmers' requirements tend to be emergent and tightly intertwined with design. Given this, many design approaches that have been targeted toward end-user programmers aim to support evolutionary or exploratory approaches. DENIM is one example (a sketching system for designing web site), which allows users to leave parts of the design in a rough and ambiguous state until that section is better understood (Newman et al 2003). The process of exploratory design can also be assisted by a design critic, which is a software feature that can review a user's design and give suggestions for improvement (Figure 10.12). For example, Janus is a tool with a visual language enabling users to design floor plans for homes (Fischer et al 1990). It contains a design critic with an extensible expert system that can identify suboptimal combinations of objects in floor plans and suggest revisions to fix those design problems. It also displays a rationale for each suggestion, so that the user can reason about whether and how to apply the advice.

Another approach is for less experienced end-user developers to seek a review from more experienced peers. By identifying short lists of best practices and providing appropriate tools, researchers have tried to make this review as efficient as possible so that it can be applied without greatly slowing the EUD lifecycle (Ronen et al 1989; Powell and Baker 2003; Rosson et al 2008). Such an approach seems most likely to be successful in an organizational setting, where end-user developers have peers that they can call upon (and where the management hierarchy can be used, if appropriate, to mandate and enforce design reviews). Another variant of this idea that has emerged recently is meta-design, a team-of-equals collaborative approach in which professional developers handle some of the design task, and end-user domain experts handle different aspects (Fischer and Giaccardi 2006; Costabile et al 2006).

Researchers have also begun to explore approaches for adapting existing software engineering design techniques to the EUD context. For example, design patterns might be relevant but need adaptation to meet the needs of end-user developers (Diaz et al 2008). Another relatively new approach is the combination of design/specifications with verification capabilities, as with the Topes system discussed earlier (Scaffidi 2009).

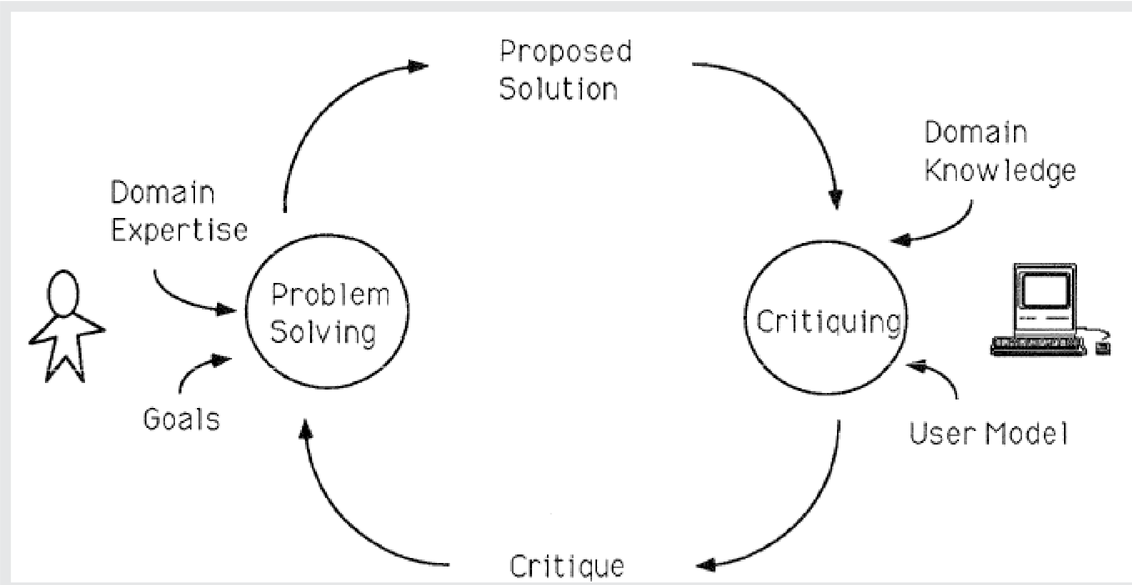



Figure 10.12: Overview of the design critic process, where the user specifies a proposed design/solution, which the design critic feature (right) reviews based on encoded domain knowledge and a model of the user's goal (Fischer et al 1990).



### 10.4.2 Verification and validation

Verification and/or validation (V&V) cover activities attempting to make sure that a program does what it is supposed to do. Testing is the most common approach for V&V (even among professional developers). One of the first works supporting V&V in EUD was to help users evaluate whether their programs contained bugs by encouraging end users to test strategically. Perhaps the most developed end-user testing approach is "What You See Is What You Test" (WYSIWYT), which guides users through the process of systematically testing spreadsheets (Fisher et al 2006). WYSIWYT employs a "Surprise-Explain-Reward" strategy (Wilson et al 2003), in which surprises such as colored borders attract users' attention to areas of the spreadsheet that need testing, and tool tips explain the colors' meaning and the potential reward in using the testing devices (Figure 10.13). Behind the scenes, WYSIWYT uses a formal test adequacy criterion to reason about elements of the formulas that have been covered by tests so far.

An alternative approach for finding errors in programs is for the programming tool to automatically look for errors on the basis of types, dimensions, or units (Erwig and Burnett 2002; Abraham and Erwig 2007; Coblenz et al 2005; Chambers and Erwig 2009). This approach can be regarded as specific kinds of assertions. For example, one system associates types with spreadsheet cells (based on the placement of labels at the top of columns and at the left end of rows) and specifies how these types propagate through the spreadsheet (Figure 10.14). If two cells with different types are combined, then their type is generalized if an applicable type is available (e.g.: "3 apples + 3 oranges = 6 fruit"), or else an error message is shown.

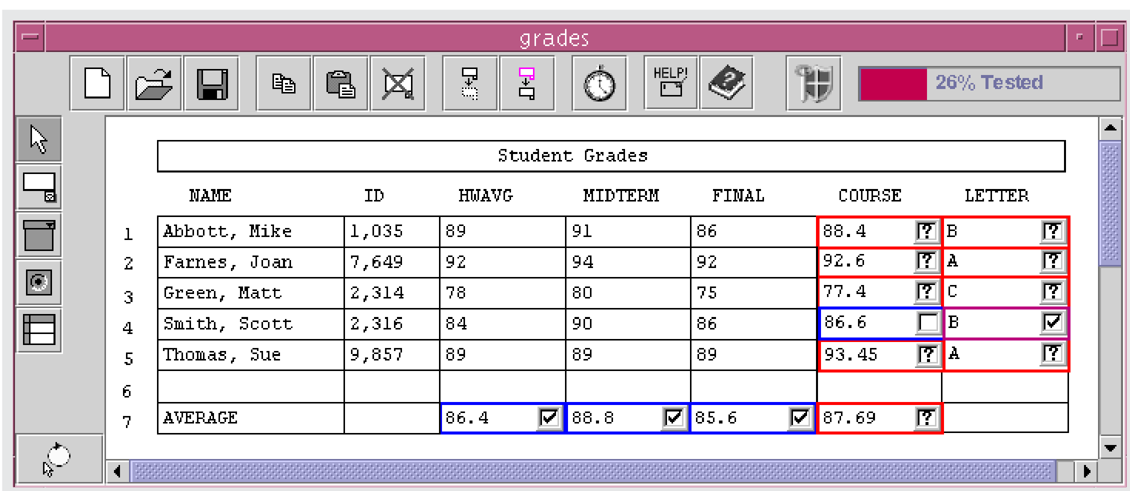



FIGURE 10.13: WYSIWYT approach, where checkmarks indicate testedness, question marks indicate that a cell needs testing, and colored borders indicate correctness.

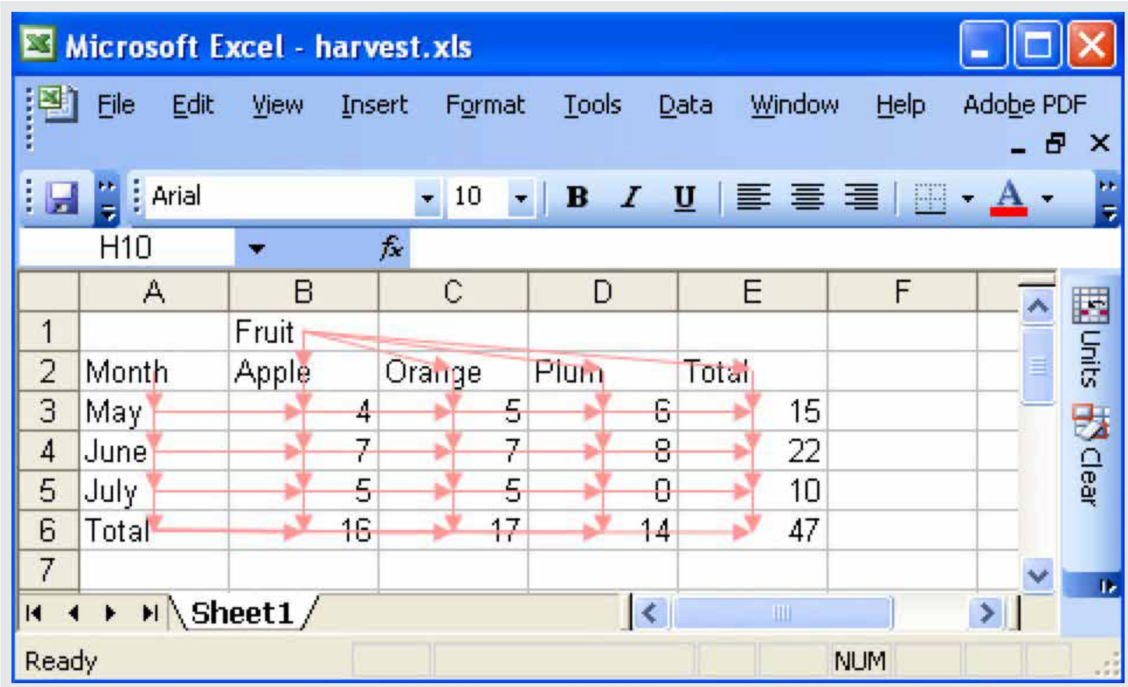



FIGURE 10.14: Using the UCheck feature to test for unit errors in a spreadsheet.



### 10.4.3 Debugging

After a programming error is detected, the next step is to remove it by debugging. Some of the debugging techniques used by professional developers have been adapted for use in EUP tools. In addition to inserting "print" statements that display the value of variables as a program executes, end user developers can step through instructions one at a time, watching for incorrect operations (Leshed et al 2008). Assertions represent another important traditional technique that has been adapted for use in EUP: a user can insert a conditional into the code, and the program will call attention to that point if the conditional evaluates to false at execution (Burnett et al 2003; Koesnandar et al. 2008; Scaffidi 2009).

Several EUP tools provide tight integration between testing and debugging. For example, assertions can be inserted proactively when a program is created, in order to perform automatic tests and initiation of debugging if an assertion fails. For instance, when a web macro is initially created, it might perform properly; however, upon later execution, invalid outputs might arise either because of a bug in the macro itself or because of changes in the structure and content of websites (the macro inputs). An assertion can catch such errors that arise, halt execution, and bring them to the user's attention to prevent the macro from running awry (Figure 10.15). Several other EUP tools that support testing techniques, such as those mentioned above, also leverage test results to facilitate debugging. For example, once incorrect spreadsheet cells have been identified through testing, dependencies can be traced back automatically to identify and highlight the formulas that are most likely to have caused those erroneous outputs (Ayalew and Mittermeir 2003; Burnett et al 2003).

A new class of debugging tools based on *question asking* has recently emerged and has proven effective in EUSE. The first tool to take this approach was the Whyline, which was prototyped for the Alice programming environment that enables users to program animations (Ko and Myers 2004). Users execute their program, and when they see a behavior they have a question about, they press a "Why" button. This brings up a menu of "why did" and "why didn't" questions, organized according to the structure of the visible 3D objects manipulated by the program. Once the user selects a question, the system analyzes the program's execution history and generates an answer explaining the error in terms of the events that occurred during execution. The Whyline approach has also been applied to debugging other kinds of programs (e.g, Ko and Myers 2008; Kulesza et al 2009).

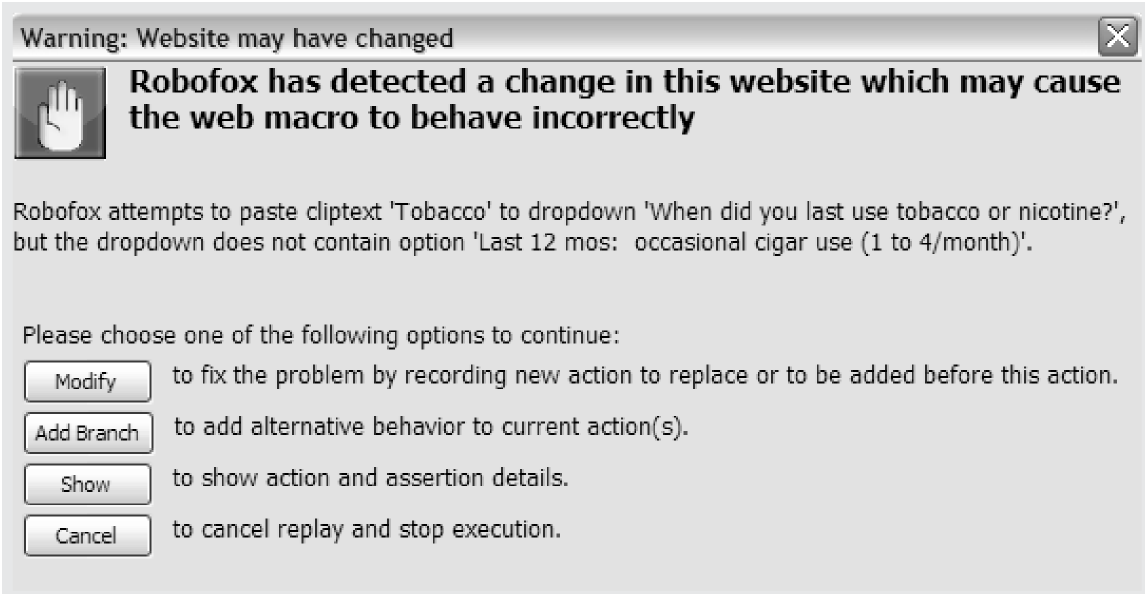



Figure 10.15: Popup window asking user to indicate whether and how a Robofox web macro should be modified due to a violated assertion.



### 10.4.4 Reuse

After code is written, reuse can speed the creation of later programs. Supporting reuse of end-user programs is challenging because end-user developers rarely have the opportunity or training required to design highly reusable programs. Another challenge is that end-user developers can make mistakes when creating programs or other files for tailoring applications, and reusing these can propagate errors across an organization (Mackay 1990). Therefore, even though systems such as repositories or file servers can make it easy for end-user developers to post programs for others to reuse, it can be extremely difficult and time-consuming for other developers to evaluate the reusability of these programs. To help reduce the difficulty of reusing programs, models of what makes end-user programs reusable are now being developed in the hopes of helping users to search repositories for reusable programs related to the user's particular interests (Scaffidi et al 2010). Outside of repositories, other work has begun to explore how to help users to extract reusable pieces (Oney and Myers 2009).

## 10.5 THE FUTURE AND IMPLICATIONS OF EUD

As users continue to grow in number and diversity, EUD is likely to play an increasingly central role in shaping software to meet the broad, varied, rapidly changing needs of the world. Along the way, further research is needed to help end-user developers create and adapt new kinds of programs in new kinds of ways. For example, as the Web 2.0 era unfolds, researchers are investigating new ways of helping users to automate the synthesis of data from multiple websites through web macros and mashups (Scaffidi et al 2008; Zang et al 2008). Another ongoing shift is the rapidly-increasing role of small mobile computers, such as smart phones; work has recently begun on enabling end users to create "apps" or other programs for these devices (Google App Inventor 2010).

With the continually broadening scope and power of end-user programming, substantial additional attention to quality will become increasingly crucial. In particular, as users continue to interact with larger numbers of anonymous peers (e.g, through social networks or "app stores"), their code may become more visible to others and therefore more susceptible to attack. Moreover, since users can now share their programs with any person on the web, many more people could be affected by an error in an end-user programmer's code. Therefore, further research is needed to help end-user developers produce software with stronger guarantees of security and privacy, without interfering with the lightweight, iterative nature of the EUD lifecycle. Moreover, as massive amounts of data become accessible through the web to users, they may need better support for designing and implementing programs with increased scalability. It will be necessary for researchers to develop new approaches, since approaches used by professional software developers, such as buffer-overrun analysis for security or Big-O analysis for performance, may be irrelevant or too complex for the needs of end-user programmers.

More broadly, EUD's continuing development as a social phenomenon has important implications for the relationship between end users and professional software developers (Fischer and Giaccardi 2006, Costabile et al 2006). The rise of EUD to date enables end users to respond to professional developers' backlog of software work, and to the reality that professional software developers are not likely to understand and plan for every user requirement when developing software. With continuing advances in EUSE, end users will not only be able to create a variety of software on their own, but they will also to be able to assess and improve that software's quality on their own—so that they know to what extent to rely upon it, and what to do to increase the software's quality if needed. As a result, the fit between software's form and individual users' needs might be closer than has been possible before, vastly increasing the usefulness of software in peoples' lives.

## 10.6 WHERE TO LEARN MORE

### 10.6.1 Books

Lieberman, H., Paterno, F., and Wulf, V. (eds.) 2006. *End User Development*, Springer.

Cypher, A, and Halbert, D. (eds.) 1993. *Watch What I Do: Programming By Demonstration*, The MIT Press.

Lieberman, H. (ed.) 2001. *Your Wish Is My Command: Programming By Example*, Morgan Kaufmann.

## 10.6.2 Relevant conferences and workshops

### CHI - Human Factors in Computing Systems

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 2011 | 2010 | 2009 | 2008 | 2007 | 2006 | 2005 | 2004 |
| 2003 | 2002 | 2001 | 2000 | 1999 | 1998 | 1997 | 1996 |
| 1995 | 1994 | 1993 | 1992 | 1991 | 1990 | 1989 | 1988 |
| 1987 | 1986 | 1985 | 1983 | 1982 | | | |

### VL-HCC - Symposium on Visual Languages and Human Centric Computing

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 2008 | 2007 | 2007 | 2006 | 2005 | 2004 | 2003 | 2002 |
| 2001 | 2000 | 1999 | 1998 | 1997 | 1996 | 1995 | 1994 |
| 1993 | 1992 | 1991 | 1990 | | | | |

Next conference is coming up 15 Sep 2013 in San Jose, CA, USA

### ISEUD - International Symposium on End User Development

2009

- Workshops on End-user Software Engineering (WEUSE I, WEUSE II, and WEUSE IV)
- ACM/IEEE International Conference on Software Engineering (ICSE)

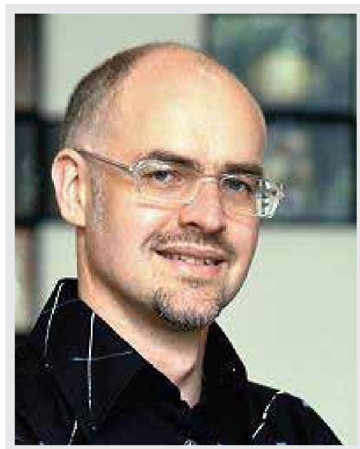
## 10.6.3 Relevant organizations

- End-Users Shaping Effective Software (EUSES) consortium
- European Spreadsheet Risks Interest Group (EUSPRIG)

## 10.8 COMMENTARY BY ALAN BLACKWELL

How to cite this commentary in your report

### Alan Blackwell



© *Alan Blackwell*

I only have one big research question, but I attack it from a lot of different angles. The question is representation. How do people make, see and use things that carry meaning? The angles from which I attack my question include various ways in which representations are applied (including design processes, interacting with technology, computer programming, visualisation), vario...

Alan Blackwell
Alan Blackwell is a member of The Interaction Design Foundation

Margaret Burnett and Chris Scaffidi provide an excellent overview of the evolving requirements in end-user development (EUD), and the technical approaches that have been taken to address them. In some ways, EUD provides one of the greatest challenges in interaction design - not because EUD tools are necessarily more complex than other specialist software categories, but because it can be harder to recognise the usability requirements. Designers in other fields must always stay alert to any presumption that customers will have exactly the same needs and habits as the designers themselves. However in the case of programming tools, it is all too easy for professional software developers to assume that every other person should approach his or her programming work in the same way as a professional does. Specialist research attention to end-users is necessary to avoid this trap when software professionals create new EUD technologies.

Successful EUD research can also provide benefits far beyond the target users. Software development can suffer from a "cobbler's children" effect - the usability of development environments is often poor by comparison to mass-market software categories where more has been invested in user research and interaction design. If EUD research leads to usability innovations for end-users, those same innovations can be later imported into mainstream software tools. A classic example is the Smalltalk language, which had superior usability because it had originally been motivated by the desire to create a programming language for children. Although children did find it useful, Smalltalk ultimately delivered far greater benefits by becoming one of the most usable object-oriented programming environments.

EUD is also providing new insights for HCI more generally. A recent focus on user-experience, and on products that meet market demand for "walk up and use" interfaces, has resulted in a lack of methods to understand requirements of more complex tasks, which involve interaction with information structures [1]. Simple media appliances and game-like interaction are welcome and desirable, but as Burnett and Scaffidi note in their conclusion, more sophisticated capabilities can also empower users. Lack of usability is frustrating, even in situations as commonplace as configuring social network privacy, or modifying a financial budget. The information structure in such tasks can be described as "programming-like", with more insight coming from EUD research than from standard usability methods. The field of Psychology of Programming [2] is able to provide guidance both to EUD, and to more domestic programming-like tasks, for example through the Cognitive Dimensions of Notations framework [3], and the Attention Investment model of abstraction use [4]. That field offers complementary insights to the very useful technical overview provided in this article.

## References and links

1. Blackwell, A.F. & Fincher, S. (2010). PUX: Patterns of User Experience. interactions 17(2), 27-31.
http://www.cl.cam.ac.uk/~afb21/publications/BlackwellFincher-PUX.pdf

2. http://www.ppig.org/

3. Blackwell, A.F. and Green, T.R.G. (2003). Notational systems - the Cognitive Dimensions of Notations framework. In J.M. Carroll (Ed.) HCI Models, Theories and Frameworks: Toward a multidisciplinary science. San Francisco: Morgan Kaufmann, 103-134.
http://www.cl.cam.ac.uk/~afb21/publications/BlackwellGreen-CDsChapter.pdf

4. Blackwell, A.F. (2002). First steps in programming: A rationale for Attention Investment models. In Proceedings of the IEEE Symposia on Human-Centric Computing Languages and Environments, pp. 2-10.
http://www.cl.cam.ac.uk/~afb21/publications/HCC02a.pdf

## 10.9 COMMENTARY BY MARY BETH ROSSON

How to cite this commentary in your report

### Mary Beth Rosson



I am a Professor of Information Sciences and Technology, and Affiliate Professor of Instructional Systems in the College of Education. With Dr. John Carroll, I co-direct the Computer-Supported Collaboration and Learning (CSCL). I am also a founding member of PSU's Center for Human-Computer Interaction, a loosely coupled university-wide consortia of researchers and educators pursuing topi...

Mary Beth Rosson
Mary Beth Rosson is a member of The Interaction Design Foundation

End-user development is happening all around us. Burnett and Scaffidi do a good job of getting this message across, as well as surveying the many paths and supports for end users who participate in software development. But one issue that is hard to pin down is the *why* and *who* of EUD. HCI researchers have known for many years that people rarely become experts even in the use of their existing software applications (Rosson, 1984). So why would they choose to develop new software?

The answer to why lies in people's real world goals. In some situations the task at hand depends entirely on building a computational artifact – e.g., creating a spreadsheet model, building a website for oneself or a group, learning science via a simulation, or connecting a set of equipment and their inputs and outputs for a lab experiment. Such situations qualify as EUD because the creator is not trained in software development and is building an artifact to meet other primary goals (e.g., education, business models). End users in these situations need tools and languages that can guide *design*: they must decompose their problems in a way that lets them map from domain knowledge and goals to computational abstractions and procedures. Sometimes a domain-specific language is the best approach, other times a generic creativity medium like a spreadsheet or web authoring tool.

An even larger set of EUD situations falls under the concept of *tailoring* as summarized by Burnett and Scaffidi. In these cases a person becomes dissatisfied with the functionality of a piece of software and changes it. This may be a change to how a bit of functionality is accessed or viewed (e.g., customization of keys, buttons or layout); a small and focused extension (e.g., defining a new document formatting style or recording steps to be repeated as a macro); or the redefinition of central aspects of an application (e.g., designing a better install configuration for a work group). Morch (1997) posits three levels of tailoring: customizing existing functionality; integrating functionality available elsewhere; and extending a system with new functionality created by end users. The degree of modification increases at each level, along with the impact of the sociotechnical context of use. That is, while simple customization may be entirely up to an individual user and his or her preferences, extending an organizational tool is the result of a complex social network of stakeholders, even if the actual change is enacted by a single individual (Mackay, 1990; Nardi, 1993). When a situation calls for tailoring, a critical aspect of the supporting EUD tools is integration with the abstractions and mechanisms of the existing application and associated practices (Costabile et al., 2006).

That leaves us with the question of who. Many scholars have pointed to the emergence of specific individuals in organizations, using metaphors like "gardener" to describe some people's willingness to synthesize, abstract and support the computational needs present in their sociotechnical context (Mackay, 1990; Nardi, 1993). Others have noted the personal characteristics of individuals who appear intrinsically motivated to notice and explore new technology options (Rosson, Ballin & Nash, 2004; Zang & Rosson, 2010). The design of EUD languages and tools might certainly begin with an emphasis on evoking and supporting these tendencies. More worrying however, is evidence that some groups of end users (e.g., women and minorities) have a sort of learned avoidance about EUD, likely due to a myriad of personal, technical and social factors (Beckwith, 2007). The paradox of the active user (Carroll & Rosson, 1987) and the attention investment model (Blackwell, 2001) offer an essential lesson: only when EUD tools or languages can suggest or evoke benefits that are perceived to be significant enough – at the moment in time that an opportunity arises – should we expect to see a full range of end users enjoying the power that such tools bring.

## References


- Beckwith, L. 2007. *Gender HCI Issues in End-User Programming*. Ph.D. Dissertation, Computer Science Department, Oregon State University, Corvallis, OR.
- Blackwell, A.F. 2002. First steps in programming: A rationale for attention investment models. In *Proceedings of the IEEE Symposia on Human-Centric Computing Languages and Environments: VL/HCC 2002* (pp. 2-10). New York, NY: IEEE Computing Society.
- Carroll, J. M., & Rosson, M. B. 1987. The paradox of the active user. In J.M. Carroll (Ed.), *Interfacing Thought: Cognitive Aspects of Human-Computer Interaction*. Cambridge, Mass: MIT Press (pp. 80-111).

- Costabile, M.F., Piccino, A., Foglio, D. & Marcante, A. 2006. Supporting interaction and co-evolution of users and systems. *Proceedings of Advanced Visual Interfaces: AVI 2006* (pp. 143-150). New York, NY: ACM Press.
- Mackay, W. 1990. *Users and Customizable Software: A Co-Adaptive Phenomenon*. Ph.D. Dissertation, Computer Science Department, Massachusetts Institute of Technology, Cambridge, MA.
- Mørch, A. 1997. Three levels of end-user tailoring: Customization, integration and extension. In M. Kyng & L. Mathiassen (Eds.), *Computers and Context* (pp. 51-76). Cambridge, MA: MIT Press.
- Nardi, B. 1993. *A Small Matter of Programming*. Cambridge, MA: MIT Press.
- Rosson, M. B. 1984. Effects of experience on learning, using, and evaluating a text-editor. *Human Factors, 26*(4), 463-475.
- Rosson, M.B., Ballin, J., & Nash, H. 2004. Everyday programming: Challenges and opportunities for informal web development. *Visual Languages and Human-Centric Computing 2004* (pp. 123-130). New York: IEEE.
- Zang, N. & Rosson, M.B. 2010. Mashups for Web-active end users. In A. Cypher, M. Dontcheva, T. Lau & J. Nichols (Eds.), *No Code Required: Giving Users Tools to Transform the Web* (pp. 409-423). San Francisco, Morgan Kaufmann.

## 10.7 BEHIND THE SCENES


*Copyright statu: Unknown (pending investigation). See section "Exceptions" in the copyright terms.*

*Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.*



## 10.10 REFERENCES

*Abraham, Robin and Erwig, Martin (2007): UCheck: A spreadsheet type checker for end users. In J. Vis. Lang. Comput., 18 (1) pp. 71-95*

*Abraham, Robin and Erwig, Martin (2004): Header and Unit Inference for Spreadsheets Through Spatial Analyses. In: VL-HCC 2004 - IEEE Symposium on Visual Languages and Human-Centric Computing 26-29 September, 2004, Rome, Italy. pp. 165-172*

*Ayalew, Yirsaw and Mittermeir, Roland (2008): Spreadsheet Debugging. In CoRR, 0*

*Ayalew, Y. and Mittermeir, R. (2003): Spreadsheet debugging. In: Proceedings of the European Spreadsheet Risks Interest Group 2003. pp. 22-23*

*Blackwell, Alan (2006): Gender in Domestic Programming: from Bricolage to Séances d'Essayage. In: CHI2006 Workshop on End User Software Engineering 2006.*

*Blackwell, Alan (2004): End-user developers at home. In Communications of the ACM, 47 (9) pp. 65-66*

*Boal, Augusto (1992): Games for Actors and Non-Actors. Routledge*

*Boehner, Kirsten, dePaula, Rogerio, Dourish, Paul and Sengers, Phoebe (2007): How emotion is made and measured. In International Journal of Human-Computer Studies, 65 (4) pp. 275-291*

*Boehner, Kirsten, dePaula, Rogerio, Dourish, Paul and Sengers, Phoebe (2005): Affect: from information to interaction. In: Bertelsen, Olav W., Bouvin, Niels Olof, Krogh, Peter Gall and Kyng, Morten (eds.) Proceedings of the 4th Decennial Conference on Critical Computing 2005 August 20-24, 2005, Aarhus, Denmark. pp. 59-68*

*Brancheau, James C. and Wetherbe, James C. (1987): Key Issues in Information Systems Management. In MIS Quarterly, 11 (1) pp. 23-45*

*Brancheau, James C., Janz, Brian D. and Wetherbe, James C. (1996): Key Issues in Information Systems Management: 1994-95 SIM Delphi Results. In MIS Quarterly, 20 (2) pp. 225-242*

*Bricklin, D., Frankston, B. and Fylstra, D. (1979). VisiCalc, Software Arts. Retrieved 31 May 2011 from http://www.bricklin.com/history/intro.htm*

*Burnett, Margaret M. (2010): End-User Software Engineering and Why it Matters. In JOEUC, 22 (1) pp. 1-22*

*Burnett, Margaret M., Cook, Curtis R., Pendse, Omkar, Rothermel, Gregg, Summet, Jay and Wallace, Chris S. (2003): End-User Software Engineering with Assertions in the Spreadsheet Paradigm. In: Proceedings of the 25th International Conference on Software Engineering May 3-10, 2003, Portland, Oregon, USA. pp. 93-105*

*Burnett, Margaret M., Sheretov, Andrei, Ren, Bing and Rothermel, Gregg (2002): Testing Homogeneous Spreadsheet Grids with the "What You See Is What You Test" Methodology. In IEEE Trans. Software Eng., 28 (6) pp. 576-594*

*Chambers, Chris and Erwig, Martin (2009): Automatic detection of dimension errors in spreadsheets. In J. Vis. Lang. Comput., 20 (4) pp. 269-283*

*Coblenz, Michael J., Ko, Andrew Jensen and Myers, Brad A. (2005): Using Objects of Measurement to Detect Spreadsheet Errors. In: VL-HCC 2005 - IEEE Symposium on Visual Languages and Human-Centric Computing 21-24 September, 2005, Dallas, TX, USA. pp. 314-316*

*Costabile, Maria Francesca, Fogli, Daniela, Mussio, Pero and Piccinno, Antonio (2006): End-user development: The software shaping workshop approach. In: Lieberman, Henry, Paterno, Fabio and Wulf, Volker (eds.). "End User Development (Human-Computer Interaction Series)". Springerpp. 183-205*

*Cypher, Allen (1993): Watch What I Do: Programming by Demonstration. Cambridge, MA, The MIT Press*

*Diaz, Paloma, Aedo, Ignacio and Rosson, Mary Beth (2008): Visual representation of web design patterns for end-users. In: Levialdi, Stefano (ed.) AVI 2008 - Proceedings of the working conference on Advanced Visual Interfaces May 28-30, 2008, Napoli, Italy. pp. 408-411*

*Dittrich, Yvonne, Lindeberg, Olle and Lundberg, Lars (2006): End-user development as adaptive maintenance. In: Lieberman, Henry, Paterno, Fabio and Wulf, Volker (eds.). "End User Development (Human-Computer Interaction Series)". Springerpp. 295-313*

*Erwig, Martin and Burnett, Margaret M. (2002): Adding Apples and Oranges. In: Krishnamurthi, Shriram and Ramakrishnan, C. R. (eds.) Practical Aspects of Declarative Languages, 4th International Symposium - PADL 2002 January 19-20, 2002, Portland, OR, USA. pp. 173-191*

*Fischer, Gerhard and Giaccardi, Elisa (2006): A framework for the future of end-user development. In: Lieberman, Henry, Paterno, Fabio and Wulf, Volker (eds.). "End User Development (Human-Computer Interaction Series)". Springerpp. 427-457*

Fischer, Gerhard, Lemke, Andreas C., Mastaglio, Thomas W. and Morch, Anders (1990): Using Critics to Empower Users. In: Carrasco, Jane and Whiteside, John (eds.) Proceedings of the ACM CHI 90 Human Factors in Computing Systems Conference 1990, Seattle, Washington,USA. pp. 337-347

Fisher, Marc, Rothermel, Gregg, Brown, Darren, Cao, Mingming, Cook, Curtis R. and Burnett, Margaret M. (2006): Integrating automated test generation into the WYSIWYT spreadsheet testing methodology. In ACM Trans. Softw. Eng. Methodol., 15 (2) pp. 150-194

Google (2010). Google App Inventor. Retrieved [Date unavailable] from Google:

Ichbiah, Daniel and Knepper, Susan (1992): The Making of Microsoft: How Bill Gates and His Team Created the World's Most Successful Software Company. Prima Lifestyles

Jones, Simon L. Peyton, Blackwell, Alan and Burnett, Margaret M. (2003): A user-centred approach to functions in Excel. In SIGPLAN Notices, 38 (9) pp. 165-176

Ko, Andrew J. and Myers, Brad A. (2004): Designing the whyline: a debugging interface for asking questions about program behavior. In: Dykstra-Erickson, Elizabeth and Tscheligi, Manfred (eds.) Proceedings of ACM CHI 2004 Conference on Human Factors in Computing Systems April 24-29, 2004, Vienna, Austria. pp. 151-158

Ko, Andrew Jensen and Myers, Brad A. (2008): Debugging reinvented: asking and answering why and why not questions about program behavior. In: Schäfer, Wilhelm, Dwyer, Matthew B. and Gruhn, Volker (eds.) 30th International Conference on Software Engineering ICSE 2008 May 10-18, 2008, Leipzig, Germany. pp. 301-310

Ko, Andrew J., Abraham, Robin, Beckwith, Laura, Blackwell, Alan, Burnett, Margaret M., Erwig, Martin, Scaffidi, Christopher, Lawrance, Joseph, Lieberman, Henry, Myers, Brad A., Rosson, Mary Beth, Rothermel, Gregg, Shaw, Mary and Wiedenbeck, Susan (2011): The State of the Art in End-User Software Engineering. In ACM Computing Surveys, 43 (3) pp. 1-44

Koesnandar, Andhy, Elbaum, Sebastian G., Rothermel, Gregg, Hochstein, Lorin, Scaffidi, Christopher and Stolee, Kathryn T. (2008): Using assertions to help end-user programmers create dependable web macros. In: Harrold, Mary Jean and Murphy, Gail C. (eds.) Proceedings of the 16th ACM SIGSOFT International Symposium on Foundations of Software Engineering November 9-14, 2008, Atlanta, Georgia, USA. pp. 124-134

Kulesza, Todd, Wong, Weng-Keen, Stumpf, Simone, Perona, Stephen, White, Rachel, Burnett, Margaret M., Oberst, Ian and Ko, Andrew J. (2009): Fixing the program my computer learned: barriers for end users, challenges for the machine. In: Proceedings of the 2009 International Conference on Intelligent User Interfaces 2009. pp. 187-196

Leshed, Gilly, Haber, Eben M., Matthews, Tara and Lau, Tessa (2008): CoScripter: automating & sharing how-to knowledge in the enterprise. In: Proceedings of ACM CHI 2008 Conference on Human Factors in Computing Systems April 5-10, 2008. pp. 1719-1728

Letondal, Catherine (2006): Participatory programming: Developing programmable bioinformatics tools for end-users. In: Lieberman, Henry, Paterno, Fabio and Wulf, Volker (eds.). "End User Development (Human-Computer Interaction Series)". Springerpp. 207-242

Lieberman, Henry, Paterno, Fabio, Klann, Markus and Wulf, Volker (2006): End-user development: An emerging paradigm. End User Development. In: Lieberman, Henry, Paterno, Fabio and Wulf, Volker (eds.). "End User Development (Human-Computer Interaction Series)". Springerpp. 1-8

Liu, Hugo and Lieberman, Henry (2005): Metafor: visualizing stories as code. In: Amant, Robert St., Riedl, John and Jameson, Anthony (eds.) Proceedings of the 2005 International Conference on Intelligent User Interfaces January 10-13, 2005, San Diego, California, USA. pp. 305-307

Mackay, Wendy E. (1990): Patterns of Sharing Customizable Software. In: Halasz, Frank (ed.) Proceedings of the 1990 ACM conference on Computer-supported cooperative work October 07 - 10, 1990, Los Angeles, California, United States. pp. 209-221

McDaniel, Richard G. and Myers, Brad A. (1999): Getting More Out of Programming-by-Demonstration. In: Altom, Mark W. and Williams, Marian G. (eds.) Proceedings of the ACM CHI 99 Human Factors in Computing Systems Conference May 15-20, 1999, Pittsburgh, Pennsylvania. pp. 442-449

Morch, Anders and Mehandjiev, Nikolay D. (2000): Tailoring as Collaboration: The Mediating Role of Multiple Representations and Application Units. In Computer Supported Cooperative Work, 9 (1) pp. 75-100

Nardi, Bonnie A. (1993): A Small Matter of Programming: Perspectives on End User Computing. Cambridge, Massachusetts, MIT Press

Newman, Mark W., Lin, James J. W., Hong, Jason I. and Landay, James A. (2003): DENIM: An Informal Web Site Design Tool Inspired by Observations of Practice. In Human-Computer Interaction, 18 (3) pp. 259-324

Oney, Stephen and Myers, Brad A. (2009): FireCrystal: Understanding interactive behaviors in dynamic web pages. In: IEEE Symposium on Visual Languages and Human-Centric Computing - VL/HCC 2009 20-24 September, 2009, Corvallis, OR, USA. pp. 105-108

Panko, R. (1998): What we know about spreadsheet errors. In Journal of End User Computing, 10 pp. 15-21

Powell, Stephen G. and Baker, Kenneth R. (2003): The Art of Modeling with Spreadsheets: Management Science, Spreadsheet Engineering, and Modeling Craft. Wiley

Repenning, Alexander and Perrone, Corrina (2000): Programming by Analogous Examples. In Communications of the ACM, 43 (3) pp. 90-97

Resnick, Mitchel, Maloney, John, Monroy-Hernández, Andrés, Rusk, Natalie, Eastmond, Evelyn, Brennan, Karen, Millner, Amon, Rosenbaum, Eric, Silver, Jay S., Silverman, Brian and Kafai, Yasmin B. (2009): Scratch: programming for all. In Communications of the ACM, 52 (11) pp. 60-67

Ronen, Boaz, Palley, Michael A. and Jr., Henry C. Lucas (1989): Spreadsheet Analysis and Design. In Communications of the ACM, 32 (1) pp. 84-93

Rosson, Mary Beth, Sinha, Hansa and Edor, Tisha (2010): Design Planning in End-User Web Development: Gender, Feature Exploration and Feelings of Success. In: Hundhausen, Christopher D., Pietriga, Emmanuel, Diaz, Paloma and Rosson, Mary Beth (eds.) IEEE Symposium on Visual Languages and Human-Centric Computing, VL/HCC 2010 21-25 September 2010, 2010, Leganés-Madrid, Spain. pp. 141-148

Rosson, Mary Beth, Carroll, John M., Seals, Cheryl D. and Lewis, Tracy L. (2002): Community design of community simulations. In: Proceedings of DIS02: Designing Interactive Systems: Processes, Practices, Methods, & Techniques 2002. pp. 75-83

Rosson, Mary Beth, Sinha, Hansa, Bhattacharya, Mithu and Zhao, Dejin (2008): Design planning by end-user web developers. In J. Vis. Lang. Comput., 19 (4) pp. 468-484

Rosson, Mary Beth, Sinha, Hansa, Bhattacharya, Mithu and Zhao, Dejin (2007): Design Planning in End-User Web Development. In: VL-HCC 2007 - IEEE Symposium on Visual Languages and Human-Centric Computing 23-27 September, 2007, Coeur dAlene, Idaho, USA. pp. 189-196

Scaffidi, Christopher (2009). Topes: Enabling End-User Programmers to Validate and Reformat Data, PhD Dissertation, Technical Report CMU-ISR-09-105. Institute for Software Research (ISR), Carnegie Mellon University

Scaffidi, Christopher, Cypher, Allen, Elbaum, Sebastian G., Koesnandar, Andhy and Myers, Brad A. (2008): Using scenario-based requirements to direct research on web macro tools. In J. Vis. Lang. Comput., 19 (4) pp. 485-498

Scaffidi, Christopher, Bogart, Christopher, Burnett, Margaret M., Cypher, Allen, Myers, Brad A. and Shaw, Mary (2010): Using traits of web macro scripts to predict reuse. In J. Vis. Lang. Comput., 21 (5) pp. 277-291

Scaffidi, Christopher, Myers, Brad A. and Shaw, Mary (2009): Fast, Accurate Creation of Data Validation Formats by End-User Developers. In: Pipek, Volkmar, Rosson, Mary Beth, Ruyter, Boris E. R. de and Wulf, Volker (eds.) End-User Development - 2nd International Symposium - IS-EUD 2009 March 2-4, 2009, Siegen, Germany. pp. 242-261

Scaffidi, Christopher, Shaw, Mary and Myers, Brad A. (2005): Estimating the Numbers of End Users and End User Programmers. In: VL-HCC 2005 - IEEE Symposium on Visual Languages and Human-Centric Computing 21-24 September, 2005, Dallas, TX, USA. pp. 207-214

Segal, Judith (2007): Some Problems of Professional End User Developers. In: VL-HCC 2007 - IEEE Symposium on Visual Languages and Human-Centric Computing 23-27 September, 2007, Coeur dAlene, Idaho, USA. pp. 111-118

Wilson, Aaron, Burnett, Margaret M., Beckwith, Laura, Granatir, Orion, Casburn, Ledah, Cook, Curtis, Durham, Mike and Rothermel, Gregg (2003): Harnessing curiosity to increase correctness in end-user programming. In: Cockton, Gilbert and Korhonen, Panu (eds.) Proceedings of the ACM CHI 2003 Human Factors in Computing Systems Conference April 5-10, 2003, Ft. Lauderdale, Florida, USA. pp. 305-312

Won, Markus, Stiemerling, Oliver and Wulf, Volker (2006): Component-based approaches to tailorable systems. In: Lieberman, Henry, Paterno, Fabio and Wulf, Volker (eds.). "End User Development (Human-Computer Interaction Series)". Springerpp. 115-141

Wulf, Volker, Pipek, Volkmar and Won, Markus (2008): Component-based tailorability: Enabling highly flexible software applications. In International Journal of Human-Computer Studies, 20 (1) pp. 1-22

Yang, Sherry, Burnett, Margaret M., DeKoven, Elyon and Zloof, Moshé M. (1997): Representation Design Benchmarks: A Design-Time Aid for VPL Navigable Static Representations. In J. Vis. Lang. Comput., 8 (5) pp. 563-599

Zang, Nan, Rosson, Mary Beth and Nasser, Vincent (2008): Mashups: who? what? why?. In: Proceedings of ACM CHI 2008 Conference on Human Factors in Computing Systems April 5-10, 2008. pp. 3171-3176

## About the authors

### Margaret M. Burnett

*© Margaret M. Burnett*

Has also published under the name of:
"Margaret Burnett"

Personal Homepage:
web.engr.oregonstate.edu/~burnett/

Current place of employment:
Oregon State University

Margaret Burnett is a Professor of Computer Science at the School of Electrical Engineering and Computer Science at Oregon State University. Her current research focuses on end-user programming, end-user software engineering, information foraging theory as applied to programming, and gender issues in those contexts. She has a long history of research in these issues and others relating to human issues of programming. She is also the principal architect of the Forms/3 and the FAR visual programming languages and, together with Gregg Rothermel, of the WYSIWYT testing methodology for end-user programmers. She was the founding project director of the EUSES Consortium, a multi-institution collaboration among Oregon State University and Carnegie Mellon, Drexel University, Pennsylvania State, University of Nebraska, University of Washington, University of Cambridge (U.K.), and IBM to help End Users Shape Effective Software.

## Christopher Scaffidi

© *Christopher Scaffidi*

Personal Homepage:
web.engr.oregonstate.edu/~cscaffid/

Christopher Scaffidi is currently an Assistant Professor of Computer Science in the School of EECS at Oregon State University. His research interests are where human-computer interaction and software engineering intersect. Most of his current projects aim to help software users to create code for themselves, and to effectively share that code with one another. He is presently serving as Director of the EUSES Consortium, an international association of seven universities and one company. He was Co-chair for the Poster/Work-in-progress track at the 2009 International Symposium on End-User Development and is on the program committee for the 2010 ACM SIGCHI Symposium on Engineering Interactive Computing Systems. He is a member of the ACM and IEEE.

## Your notes and thoughts on chapter 10

Record your notes and thoughts on this chapter. If you want to share these thoughts with others online, go to the bottom of the page at:
http://www.interaction-design.org/encyclopedia/end-user_development.html

**NOTES:**

# CHAPTER 11

# Philosophy of Interaction - and the Interactive User Experience

***by Dag Svanaes.***

Over the last two decades, *interaction design* has emerged as a design discipline alongside traditional design disciplines such as graphics design and furniture design. While it is almost tautological that furniture designers design furniture, it is less obvious what the end product of interaction design is. Löwgren's answer is "interactive products and services" (Löwgren 2008). This narrows it down, but leaves open the question of what it means for something to be interactive.

Interactive systems have been studied within the field of Human-Computer Interaction since the early 1980s. This research has given us valuable knowledge about users, systems and design methodology, but few have asked "philosophical" questions about the very nature of interactivity and the interactive user experience.

I will approach the question of interactivity from a number of angles, in the belief that a multi-paradigmatic analysis is necessary to give justice to the complexity of the phenomenon. I will start by defining the scope through some examples of interactive products and services. Next, I will analyse interactivity and the interactive user experience from a number of perspectives, including formal logic, cognitive science, phenomenology, and media and art studies. A number of other perspectives, e.g. ethnomethodology, semiotics, and activity theory, are highly relevant, but are not included here. *(For an analysis that includes these perspectives, see (Svanaes 2000)).*

Video 11.1: Introduction to Philosophy of Interaction.

VIDEO 11.2: Guiding Principles of Interaction Design derived from Heidegger.

*Courtesy of Rikke Friis Dam and Mads Soegaard. Copyright: CC-Att-ND (Creative Commons Attribution-NoDerivs 3.0 Unported). View full screen or download (960 KB)*

VIDEO 11.3: Principles of Interaction Design derived from Merleau-Ponty.

*Courtesy of Rikke Friis Dam and Mads Soegaard. Copyright: CC-Att-ND (Creative Commons Attribution-NoDerivs 3.0 Unported). View full screen or download (5 MB)*

VIDEO 11.4: Advantages and Problems with Cognitivism.



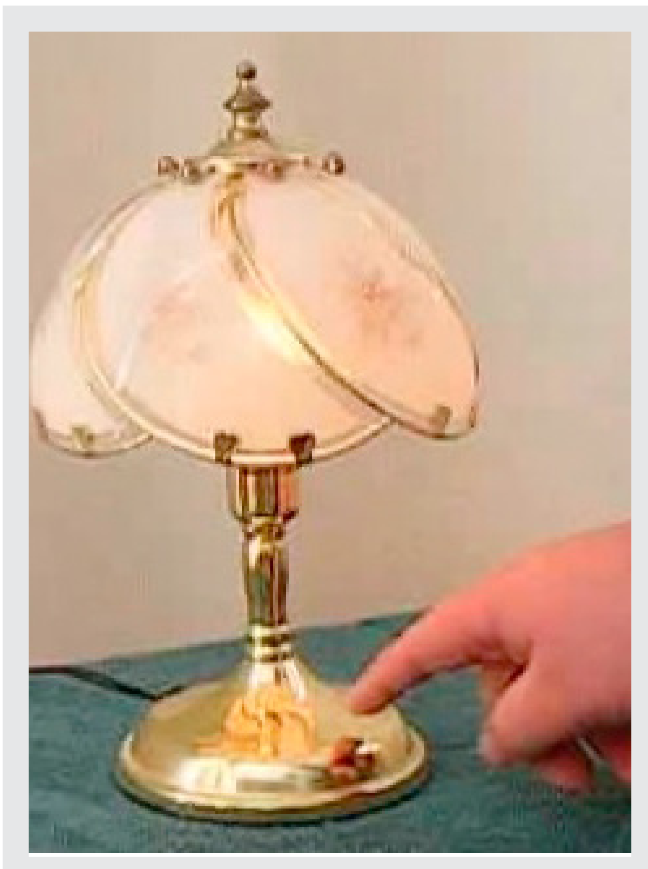
## 11.1 TERMINOLOGY

The Merriam-Webster dictionary defines *interaction* as "mutual or reciprocal action or influence". Taking this definition as a starting point, what is the meaning of *interactive* and *interactivity*? A product or service is *interactive* if it allows for interaction. An artefact's *interactivity* is its interactive behaviour as experienced by a human user. Or to be more precise, it is the potential for such experiences. Its *interactivity* is a property of that artefact; alongside other properties like its visual appearance. *Interactivity* can also be used as a noun to signify everything interactive, similar to how radioactivity refers to everything radioactive.

Many definitions exist for "the user experience". I prefer this one: "a person's perceptions and responses that result from the use or anticipated use of a product, system or service" (ISO 2009)

## 11.2 THE SCOPE: INTERACTIVE PRODUCTS AND SERVICES

What makes a product or service interactive? One of the simplest interactive products imaginable is a touch-sensitive light switch like the one in Figure 11.1A. You touch it once to turn the light on, and again to turn the light off. At the other end of the complexity scale you find interactive products like the cockpit of a modern aeroplane (Figure 11.1B); allowing trained pilots to fly the plane through a number of input devices and visual displays.



*Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.*

*Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.*

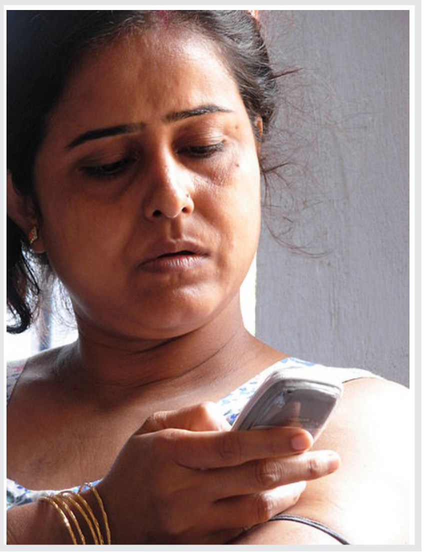
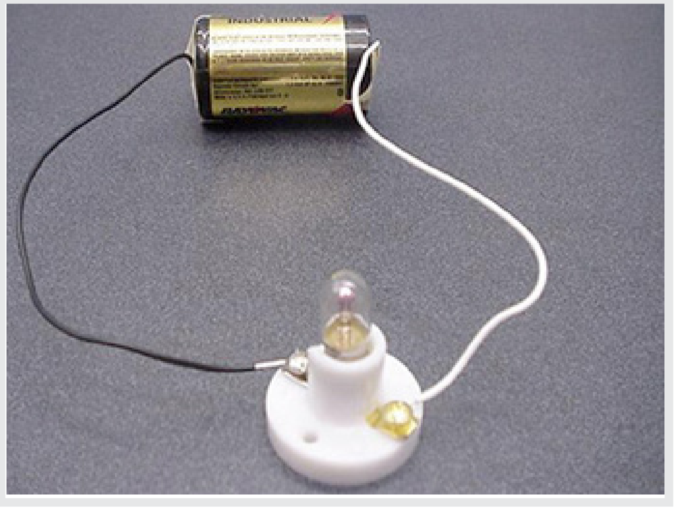




Figure 11.1 A-B-C-D: Examples of interactive products and services.

Examples of interactive services include internet banking, online shopping, and social media, all made possible through networked digital devices like PCs (11.1C) and mobile phones (11.1D). All above examples are interactive. Are there digital products that are not interactive?





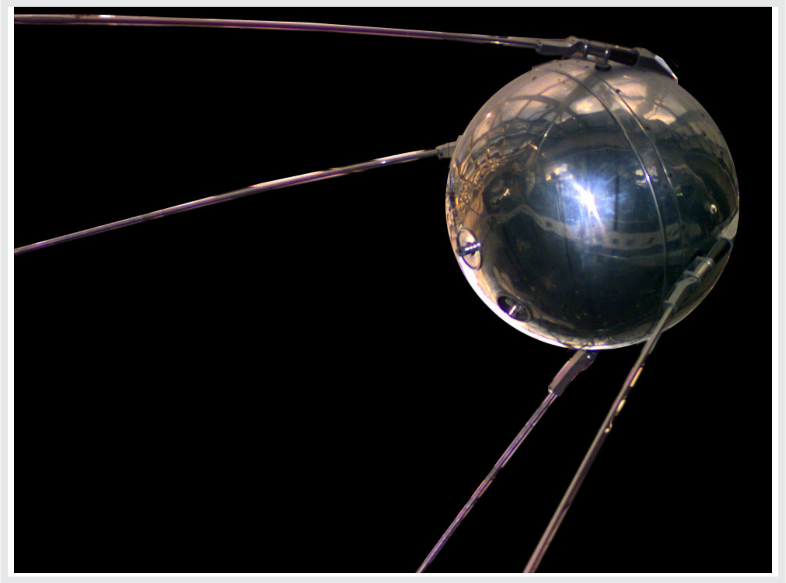




FIGURE 11.2 A-B-C: Some non-interactive artefacts.

If you solder a light bulb to a battery and leave it on your desk until the battery is drained (Figure 11.2A), this digital product can hardly be called interactive. You can of course turn it off by cutting one of its wires, but that would not be an intended interactivity of the product. The light bulb could be substituted with something far more complex, like a digital photo frame that was programmed to generate random fractals on a screen (Figure 11.2B). With no buttons, handles or other means for interacting, despite its complex behaviour, neither that product would be interactive. It would be like the 1957 Sputnik 1 satellite (Figure 11.2C), which contained a "transmit only" radio beacon that transmitted beeps from space for 20 days until its batteries ran out.

From the above examples it becomes clear that what makes a product or service interactive is not its complexity, nor the fact that it is digital, but whether it is designed to respond to actions by a user.

## 11.3 FORMAL DESCRIPTIONS OF INTERACTIVE BEHAVIOUR

One way of describing the interactive behaviour of a product or a service is through a formal representation. A number of such formalisms exist, the simplest being *state diagrams*. A state diagram is a visual representation of a Finite State Machine.

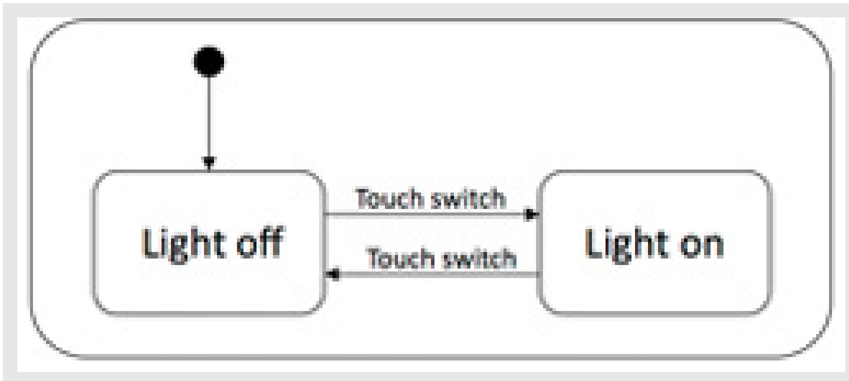
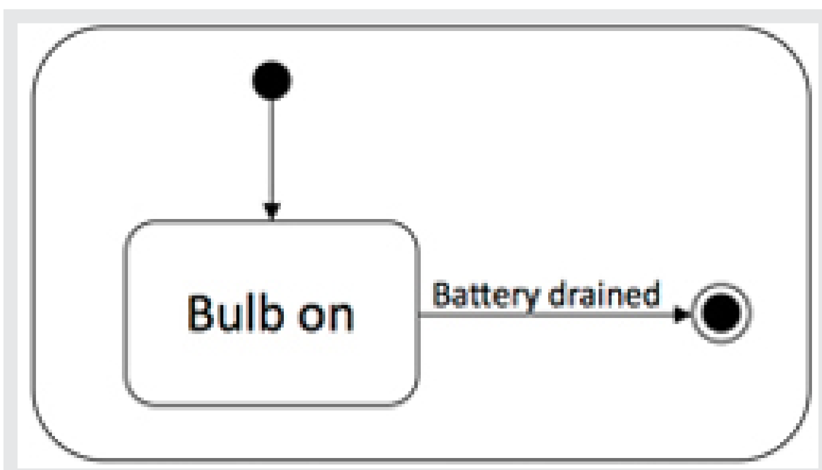



Figure 11.3A shows a state diagram for the touch-sensitive light switch in Figure 11.1A. It contains two states, "Light off" and "Light on", and two user-initiated transitions between the states ("Touch switch"). The black dot leading in to the "Light off" state tells us that this is the initial state, i.e. the light starts out in the "off" state.

*Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.*




*Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.*

Figure 11.3 A-B: State diagrams for light switch and "Bulb with battery"

In Figure 11.3B, we see the state diagram for the non-interactive "Bulb-with-battery" device. It starts with the light being on and stays in that state until the battery is drained. The black dot in a circle is the "game over" symbol. When the battery is drained, the device stops being what it was intended to be.

A number of more sophisticated formalisms have been used for describing interactive behaviour, including Harel's hierarchical state diagrams (Harel 1987), temporal logic (Hartson and Gray 1992), Petri nets (Elkoutbi and Keller 2000) and algebra (Thimbleby 2004).

Formal representations of interactive behaviour are well suited to describe the technical side of interactivity, but say little of the human side. They are of little value in answering questions like: "How is the interaction experienced?", "What does the interaction mean to the user?" To be able to answer such questions about the interactive user experience, we have to leave formal logic and the natural sciences and turn to the humanities and the social sciences.

## 11.4 COGNITIVE SCIENCE: INTERACTION AS INFORMATION PROCESSING

Since the birth of Human-Computer Interaction (HCI) as a scientific discipline in the 1980s, cognitive science has been the dominant paradigm for describing the human side of the equation. "The Psychology of Human-Computer Interaction" by Stuart K. Card, Thomas P. Moran and Allen Newell (Card et al 1983) presented a model of the user based on an information processing metaphor (Figure 11.4). Here, the interaction is modelled as information flowing from the artefact to the user, where it is processed by the user's "cognitive processor", leading to actions like pushing a button. Their model sees interaction as the sum of stimuli reception and user actions.

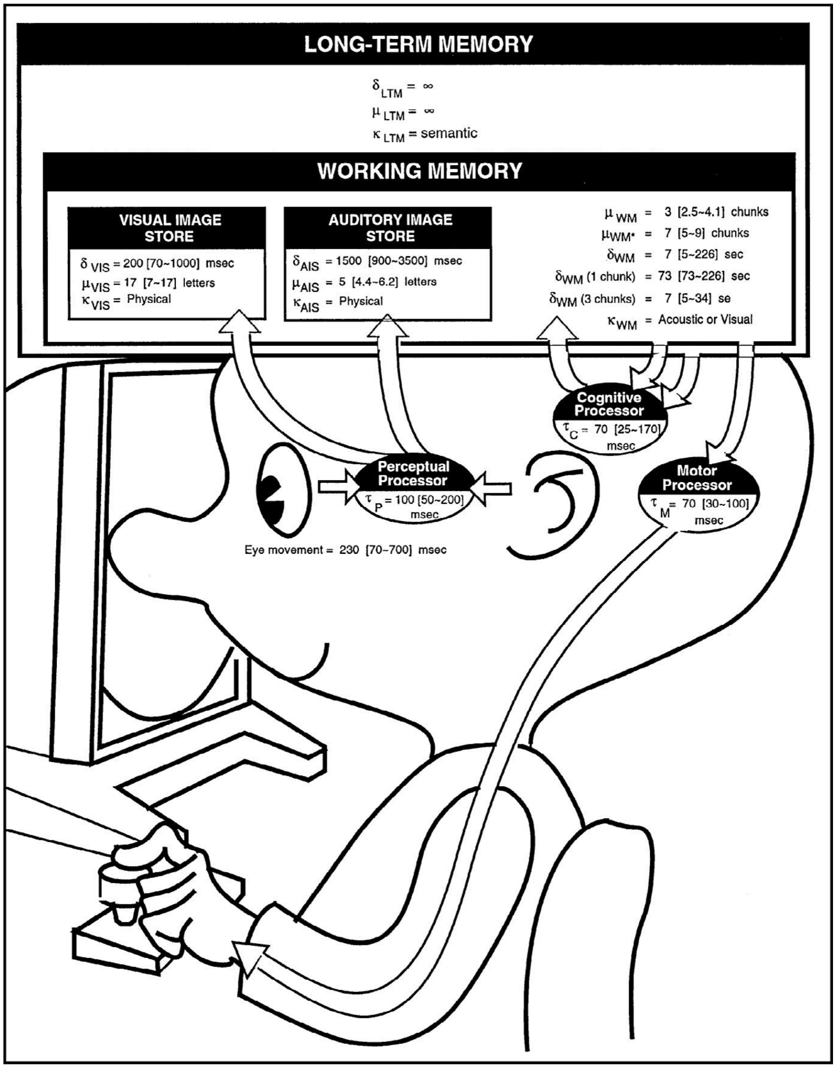



FIGURE 11.4.A: Modelling the user as an information processor.

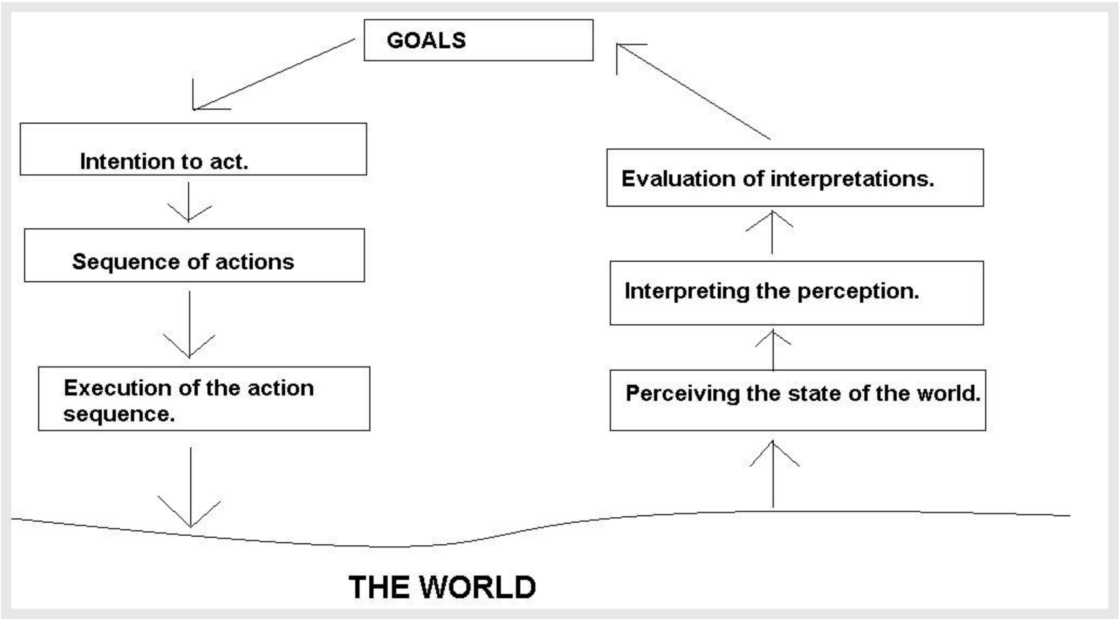



FIGURE 11.4.B: The seven step action cycle.



Imagine the user in Figure 11.4A operating the light switch in Figure 11.1A. The act of turning on the light switch would be modeled as information about the state of the light reaching the *perceptual processor* through the user's eyes, where it would flow to the *working memory*, and be processed by the *cognitive processor*. A command would then be sent to the *motor processor*, leading to the hand pushing the switch.

In "The Design of Everyday Things" (Norman 1988), Don Norman elaborates the details of what is going on as a seven step "action cycle" (Figure 11.4B). Returning to our user in front of the light switch in Figure 11.1A, Norman would describe this as the user having the goal of turning the light on (step 1). This goal would lead to an intention to act (step 2), leading to a sequence of actions being sent to the motor processor (step 3), where it would trigger a hand movement (step 4). In “the world”, the light would turn on, and this would be perceived (step 5) and interpreted (step 6) by the user. Finally, the user would evaluate the new state of the light as a fulfillment of the goal (step 7), and be ready for a new action cycle. The action cycle is described by Norman in the following video.

VIDEO 11.5: The action cycle as described by Don Norman.



Based on their model of the user, Card, Moran and Newell devised a framework for predicting user behaviour called GOMS (Goals, Operators, Methods, and Selection Rules). A number of GOMS-inspired cognitive frameworks have since been developed to model the behaviour of the user, all based on the same basic assumptions of the human information processing model.

GOMS-like models have been successful in predicting key-level human behaviour for routine tasks, but have shown little explanatory and predictive power when it comes to more open tasks, like updating your Facebook profile. Further, they are of little help in understanding the interactive user experience.

In "The Design of Everyday Things", Don Norman introduced the concept of *affordance* that had been developed by the psychologist J.J. Gibson. Norman defines affordance: "...the perceived and actual properties of the thing, primarily those fundamental properties that determine just how the thing could possibly be used". While Gibson's ecological approach to human cognition and perception in many respects is incommensurate with the information-processing approach, the *affordance* concept has mainly been interpreted within HCI to describe what functions an object allows for, and how this is "signalled" through its visual appearance. Norman illustrates the affordance concept in the video below. Adding the concept of *affordance* to the framework, the light switch in Figure 11.1A would appear to the user as an object that *affords* turning the light on and off.

Don Norman's illustration of the Affordance concept.

Video 11.6: Affordances.

A number of researchers in HCI have argued that the information-processing model reduces the user to a mechanical symbol-processing machine, leaving out important aspects of what defines us as human. One of the earliest criticisms of the information processing approach to human-computer interaction was voiced by Stanford professor Terry Winograd and Fernando Flores in their influential book "Computers and Cognition" (Winograd and Flores 1986). The book was primarily written as a criticism of artificial intelligence and cognitive science, but has strong relevance for a discussion of interactivity.

Winograd and Flores presented three alternatives to cognitive science, of which the phenomenology of the German philosopher Martin Heidegger (1889-1976) is the most relevant here.

## 11.5 HEIDEGGER: INTERACTION AS TOOL USE

Winograd and Flores argue in "Computers and Cognition" that cognitive science takes for granted that human cognition and communication are symbolic, and that symbols like "cat" refer in a one-to-one manner to objects in the world. Heidegger's philosophy of being (Heidegger 1996; original version is Heidegger 1927) rejects this view and starts out with our factual existence in the world and the way in which we cope with our physical and social environment. His philosophy spans a wide range of topics, of which Winograd and Flores mainly use his analysis of tools. Heidegger used a carpenter and his hammer as an example (Figure 11.5B). Winograd and Flores argue that a computer can be viewed as a tool: For skilled users of computers, the computer is transparent in use - it is *ready-to-hand*. When I write a document in a text editor, my focus is on the text and *not* on the text editor. If my text editor crashes, my focus is moved from the text that I am working on to the text editor itself. It is only when we have a *breakdown situation* and the computer stops working as a tool that it emerges as an object in the world - it becomes *unready-to-hand*. If we are not able to fix the problem that causes the breakdown, it becomes *present-at-hand*.







Figure 11.5 A-B-C-D: The "hammerness" of things.

Heidegger would describe the light switch in Figure 11.1A as a tool for controlling the light. As part of our everyday life, a light switch is an integral part of our background of *readiness-to-hand*, and the interaction with the switch is to some extent invisible to us. It is only when the switch stops working as expected, or when we consciously chose to reflect on it, that it emerges from the *background* as an object.

Heidegger does not deny the fact that the light switch exists in the world as an object to be viewed, touched and manipulated. His point is that the essence of the switch only emerges through use. Its "switchness" is hidden for us until we put it into use. An important aspect of its "switchness" is that it allows for a certain kind of interaction. When the ape in Kubrick's *2001: A Space Odyssey* (Figure 11.5C) realises that the piece of bone in front of him can be used to crack things, the "hammerness" of the bone emerges to him - and bones forever stop only being bones. The bone's "hammerness" had been there all the time, but it needed to be put into practice to emerge. Similarly with the interactivity of a light switch - its "switchness" emerges through use.

From a Heideggerian perspective, the specific meaning of the interaction with the light switch depends on the use situation and the user's intention. Turning the light on as part of my everyday action of entering a room is different from turning the light on to see if the switch works. In the first case the interaction is part of a wider goal, while in the second case it would be a goal in itself. Cognitive science would miss this subtle difference, as it would model both interactions as the same goal-seeking information processing behaviour. Heidegger would also argue that to be able to understand how an interaction is meaningful for a specific user, we would have to understand the *lifeworld* of that user, i.e. the cultural and personal background that serves as a frame of reference and context for every experience of that person.

Heidegger further argues that tools exist in the shared practice of a culture as part of an *equipmental nexus* , e.g. hammers with nails and wood. The hammer gets it significance through its relation to nails and wood, as the nail get its significance through its relation to hammer and wood. The elements form a whole, and each element gets its significance from its role in this whole.

## 11.6 MERLEAU-PONTY: INTERACTION AS PERCEPTION

Maurice Merleau-Ponty (1908-1961) was, besides Jean Paul Sartre, the most influential French philosopher of the 1940s and 1950s. Inspired by Heidegger, Merleau-Ponty stressed that every analysis of the human condition must start with the fact that the subject is in the world. This *being-in-the-world* is prior to both object perception and self-reflection. To Merleau-Ponty, we are not Cartesian self-knowing entities detached from external reality, but subjects already existing in the world and becoming aware of ourselves through interaction with our physical environment and with other subjects.

In his major work, "The Phenomenology of Perception" (original: Merleau-Ponty 1945; Translated: Merleau-Ponty 1962), Merleau-Ponty performs a phenomenological analysis of human perception. His purpose is to study the "precognitive" and embodied basis of human existence. He ends up rejecting most of the prevailing theories of perception at his time. In all his writing there is a focus on the first-person experience. Merleau-Ponty rejected the idea of perception as a passive reception of stimuli. When we perceive objects with our eyes, this is not a passive process of stimuli reception, but an active movement of the eyeballs in search of familiar patterns. This view is in total opposition to the popular view in "information-processing" HCI that sees perception as sense data being passively received by the brain. To Merleau-Ponty there is no perception without action; perception requires action.

Figure 11.6.A: Rapid eye movements of layperson.



Figure 11.6.B: Rapid eye movements of artist.

Perception hides for us this complex and rapid process going on “closer to the world” in “the pre-objective realm”. Modern eye trackers allow us to see these rapid perceptual interactions unfold in vision. The optics of the human eye is such that we see the world through a rapidly moving peephole. In Figure 11.6 we see the rapid eye movements of two different persons viewing the same painting, a non-artist (11.6A) and a trained artist (11.6B). We see how the artist rapidly scans the whole painting (Figure 11.6B), while the layperson mostly focuses on the face of the girl (Figure 11.6A) (from (Vogt and Magnussen 2007)). The result of their different viewing styles is that they actually see different paintings. Merleau-Ponty uses the term *phenomenal field* to denote the personal background of experiences, training, and habits that shapes the way in which we perceive the world.

Merleau-Ponty saw perception as an active process of meaning construction involving large portions of the body. The body is, a priori, the means by which we are intentionally directed towards the world. When I hold an unknown object in my hand and turn it over to view it from different angles, my intentionality is directed toward that object. My hands are automatically coordinated with the rest of my body and take part in the perception in a natural way. Any theory that locates visual perception to the eyes alone does injustice to the phenomenon. To Merleau-Ponty, the body is an undivided unity, and it is meaningless to talk about the perceptual process of seeing without reference to all the senses, to the total physical environment in which the body is situated, and to the “embodied” intentionality we always have toward the world.

The body has an ability to adapt and extend itself through external devices. Merleau-Ponty used the example of a blind man’s stick to illustrate this. When I have learned the skill of perceiving the world through the stick, the stick has ceased to exist for me as a stick and has become part of “me”. It has become part of my body and at the same time changed it.

Applied to an analysis of interactivity, Merleau-Ponty invites us to see interaction as perception. If I test out the light switch in Figure 11.1A to see if it works, this interaction can be seen as a perceptual act involving both eyes and hand. I move my hand to the switch as part of the process of perceiving its behaviour, in the same way as my eyes make rapid eye movements when I see a painting. The hand movements towards the switch result from my directedness towards the object of perception, i.e. the behaviour of the switch.

In more complex interactions, like when an experienced computer user plays World-of-Warcraft, the perceiving body extends into the game. When the gameplayer tries out a new sword that she has acquired for her game character, she perceives its working through the mouse and the part of the software that let her control her character. Playing World-of-Warcraft is similar to riding a bicycle or driving a car in that the technology becomes a tool, but it differs in that the world is computer generated.

The integrated view of action and perception makes Merleau-Ponty an interesting starting point for a discussion of meaningful interactive experiences. A consequence of his theory is that it should be possible to lead users into interactions with the computer that are meaningful at a very basic level. The interactions themselves can be meaningful.

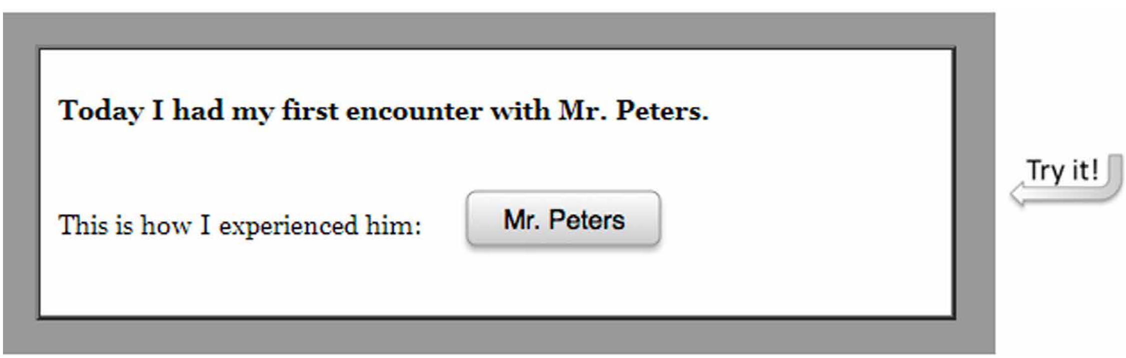



Figure 11.7: The Mr. Peters button.

The interactive artefact in Figure 11.7 exemplifies this. Try it by clicking on the "Mr. Peters" button!

The button has a script that makes it jump when the cursor is moved over it. The user tries to click on the button, but experiences that “Mr. Peters” always “escapes”. Most users understand the intended meaning of the example and describe Mr. Peters as a person who always avoids you, a person you should not trust. The interaction itself works as a metaphor for Mr. Peters’ personality. How does the philosophy of Merleau-Ponty shed light on this example?

### 11.6.1 Key points

#### 11.6.1.1 Perception requires action

- Perception of the “Mr. Peters” button requires action. The button as interactive experience is the integrated sum of its visual appearance and its behaviour. Without action, we are left with the visual appearance of the button, not the actual button as it emerges to us through interaction.

#### 11.6.1.2 Perception is an acquired skill

- One of the necessary conditions for the Mr. Peters example to work is that the user has acquired the skill of moving the mouse cursor around. This skill (Merleau-Ponty: habit) is part of being a computer user. Without this skill, the only perception of the Mr. Peters button would be its visual appearance.

#### 11.6.1.3 Tool integration and bodily space

- For the trained computer user, the mouse has similarities with the blind man’s stick. The physical mouse and the corresponding software in the computer are integrated into the experienced body of the user. The computer technology, and the skills to make use of it, changes the actual bodily space of the user by adding to the potentials for action in the physical world also the potentials for action presented by the computer. The world of objects is in a similar manner extended to include also the "objects" in the computer.

#### 11.6.1.4 Perception is embodied

- Experiencing the "Mr. Peters" button requires not only the eye, but also arm and hand. Mouse movements and eye movements are integrated parts of the perceptual process that lead up to the perception of the button's behaviour. The interactive experience is both created by and mediated through the body.

#### 11.6.1.5 Intentionality towards-the-world

- As a skilled computer user, I have a certain "directedness" towards the computer. Because of this intentionality, the Mr. Peters button presents itself to me not only as a form to be seen, but also as a potential for action with an expectation for possible reactions. From seeing the button to moving the cursor towards it, there is no need for a "mental representation" of its position and meaning. The act of trying to click on the button is part of the perceptual process of exploring the example. When the button jumps away, I follow it without having to think.

#### 11.6.1.6 The phenomenal field

- In the above example, the context of the button is given by the leading text and by the user's past experiences with graphical user interfaces. It is important to notice that this example only works with users who are used to clicking on buttons to find more information. This is the horizon of the user, i.e. the phenomenal field that all interaction happens within. The Mr. Peters button emerges

> as a meaningful entity because the appearance of a button on a computer screen leads to a certain expectation and a corresponding action. The action is interrupted in a way that creates an interactive experience that is similar to that of interacting with a person who always escapes you.

With Merleau-Ponty it becomes meaningless to talk about interaction as the sum of stimuli reception and action as cognitive science tells us. Interaction is better described as a kind of perception. I perceive the behaviour of the "Mr. Peters" button through interaction. This perception involves both hand and eye in an integrated manner. Interaction-perception is immediate and "close to the world".

## 11.7 A MEDIA AND ART PERSPECTIVE

While phenomenology can help us understand the interactive user experience for a specific product, and might help us choose between two or more alternative designs, it gives us little guidance on what designs are possible. To be able to fully utilize the potential of interactive media, it is important to have a deep understanding of the medium itself. There is a tradition in Media and Art Studies for asking questions concerning the nature of the medium being studied. However, compared to the vast literature on the social and cultural impact of new media, media studies with a focus on the properties of the medium itself are rare. The most prominent author on this subject is Rudolf Arnheim (1904-2007). Arnheim dealt with non-interactive media like film, painting, drawing, sculpture, and architecture, and he analysed their media-specific properties from an artistic and psychological perspective. In the introduction of "Art and Visual Perception: a Psychology of the Creative Eye" (Arnheim 1974), he states explicitly that he is not concerned with the cognitive, social, or motivational aspects. Nor is he concerned with "the psychology of the consumer". By ignoring all elements of social function and meaning in a traditional sense, he was free to discuss issues such as balance, space, shape, form, and movement in relation to the different media. Arnheim draws heavily on examples from art and gestalt psychology. What is relevant for the current study is not his results, but his approach to the study of a new medium.

We find a similar approach to studying a medium in artist and Bauhaus teacher Johannes Itten's theory of colours.

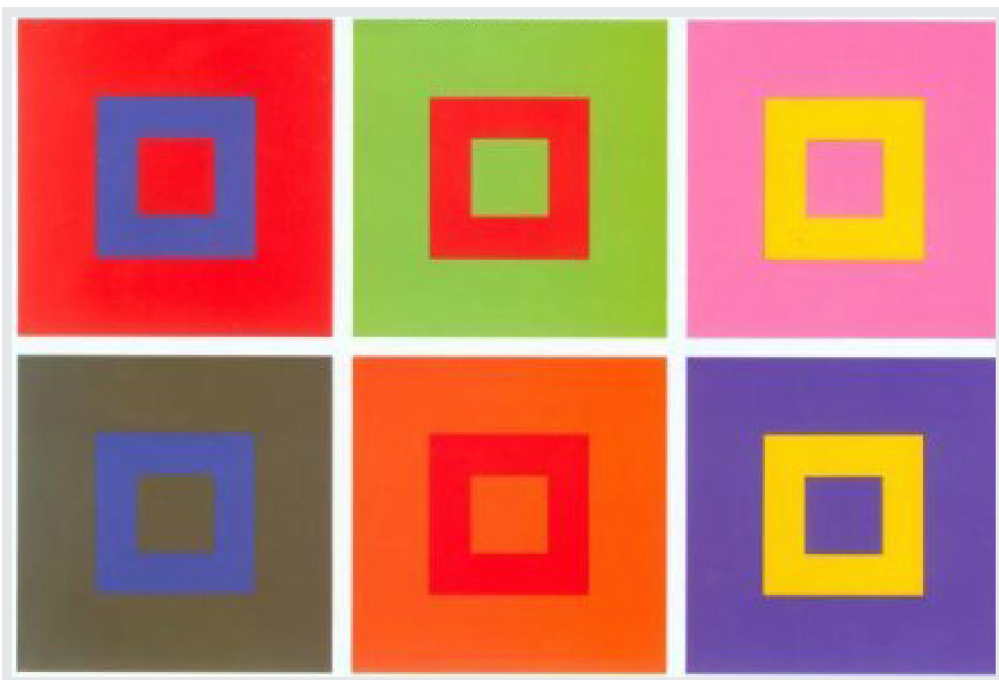


FIGURE 11.8: One of Itten's explorations of the interplay between colors.



In Figure 11.8 we see some of the coloured squares that he drew to illustrate how the perception of colour changes with the background (Itten 1974).

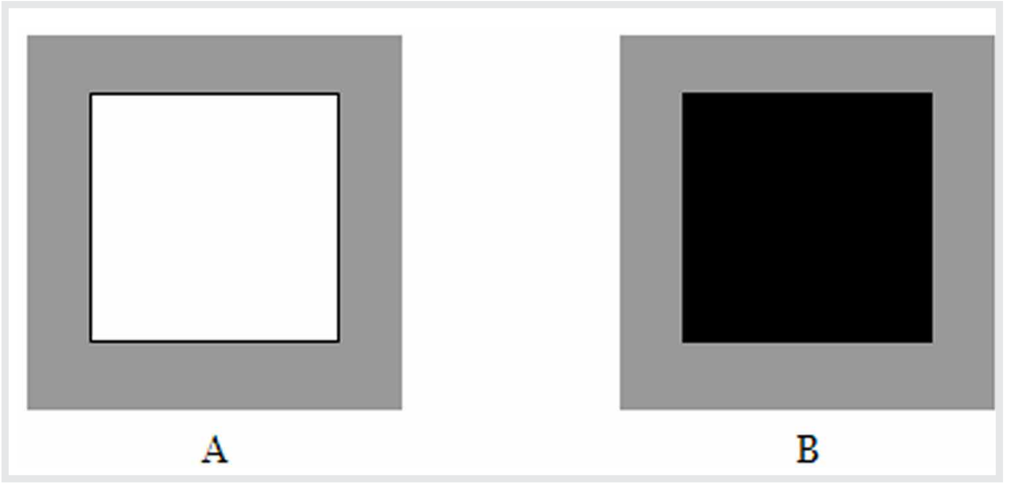



Figure 11.9: A white and a black pixel enlarged.

Itten brought to the Bauhaus School of Design the idea that students of design should develop a deep knowledge of their media and materials through explorations of their properties. Seen as a medium and a material, the modern computer can be viewed as a display of pixels that each can have only one colour at any given time. Through some input device(s), the user can interact with this matrix of pixels. In Figure 11.9 we see a white and a black pixel enlarged.

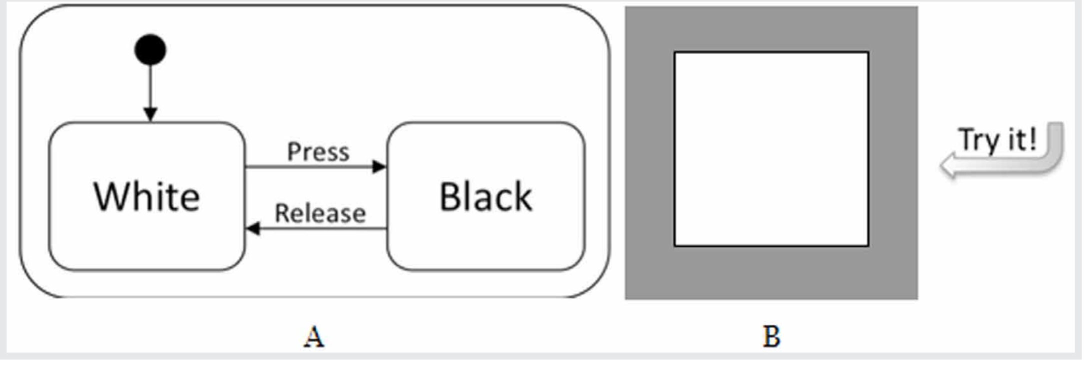



Figure 11.10: An interactive pixel with corresponding state diagram.

The two pixels in Figure 11.9 are static in the sense that they do not respond to input from the user. An interactive pixel is a pixel that responds to user actions. The simplest "artefact" of this kind is a pixel that changes colour when clicked on. The pixel in Figure 11.10B has what we call "push button" behaviour. Figure 11.10A shows its state diagram.


Figure 11.11: Two interactive pixels that only differ in behaviour.

The two interactive artefacts in Figure 11.11 look the same, but differ in behaviour. The "user experience" of an interactive artefact is the sum of its visual appearance and its interactive behaviour. The behaviour can only be experienced through interaction, and requires an active user. The fact that the pixel in Figure 11.11A is a "push button" and the pixel in Figure 11.11B is a "toggle" can only be perceived through interaction.

Borrowing from gestalt psychology, I use the term *interaction gestalts* for these kinds of basic interactive user experiences (Svanaes 1993).

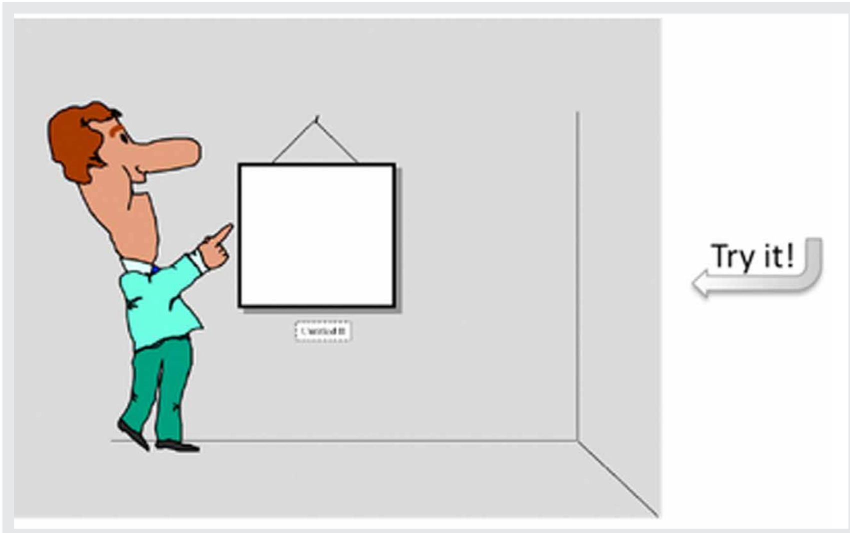



Figure 11.12: Abstract interactive art. It starts out white and goes black when it is touched.

If Figure 11.11A were placed in an art gallery, such an artefact would require the interactive behaviour to be perceived through use. Figure 11.12 illustrated this. A detached observer would miss the essence of this piece of minimalist interactive abstract art.

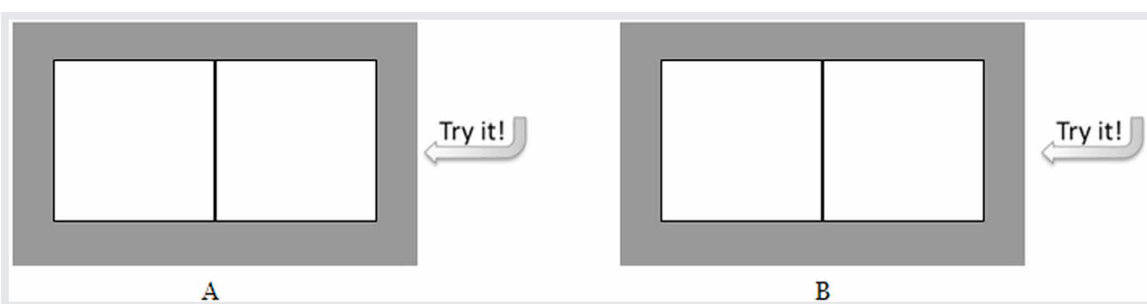



Figure 11.13: Two two-pixel interactive artefacts.

The artefacts in Figure 11.13 each consist of two pixels. Their behaviours are so simple that we can take them in as wholes, i.e. as *interaction gestalts*.

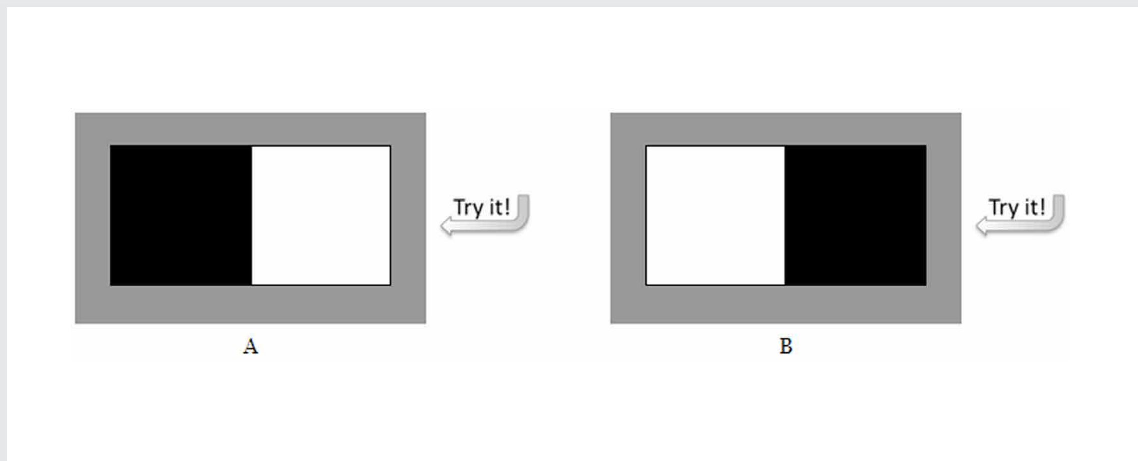



FIGURE 11.14: Spatialisation: black on white (A) and white on black (B).

The previous artefacts all had the response located to the pixel you click on. By allowing pixels to affect each other, we get more complex artefacts. The artefacts in Figure 11.14 illustrate this. When clicked on, Figure 11.14A creates a foreground and a background; a black square moving on a white background. In Figure 11.14B the white square is active. This creates the illusion of a white square moving on a black background.

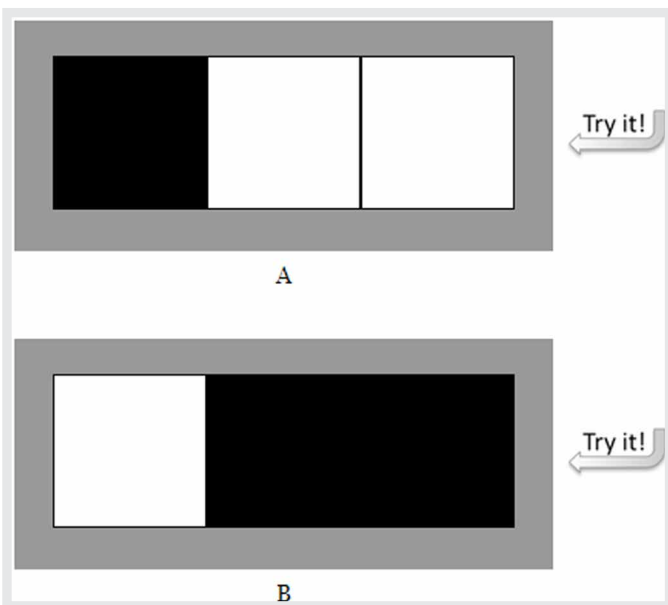



FIGURE 11.15: Spatialisations with three pixels.

With three pixels like in Figure 11.15, the spatialisation becomes clearer. This simple swapping of colours is what happens at the pixel level when objects move on the computer screen.

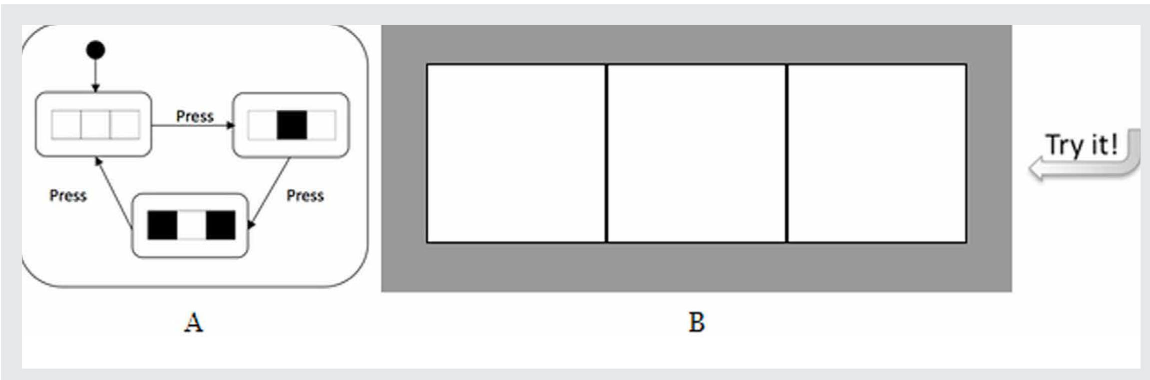



FIGURE 11.16: State-space example.

The three-pixel artefact in Figure 11.16B has three states. As the artefact does not create the illusion of foreground/background, we perceive the behaviour as a rotation between states. It starts out all white, and after two states we are back "home" where we started. Figure 11.16A shows its state diagram. This kind of "state space" is the perceptual basis for the World-Wide-Web metaphor of moving between web pages, each web page being a state of the screen's pixels.

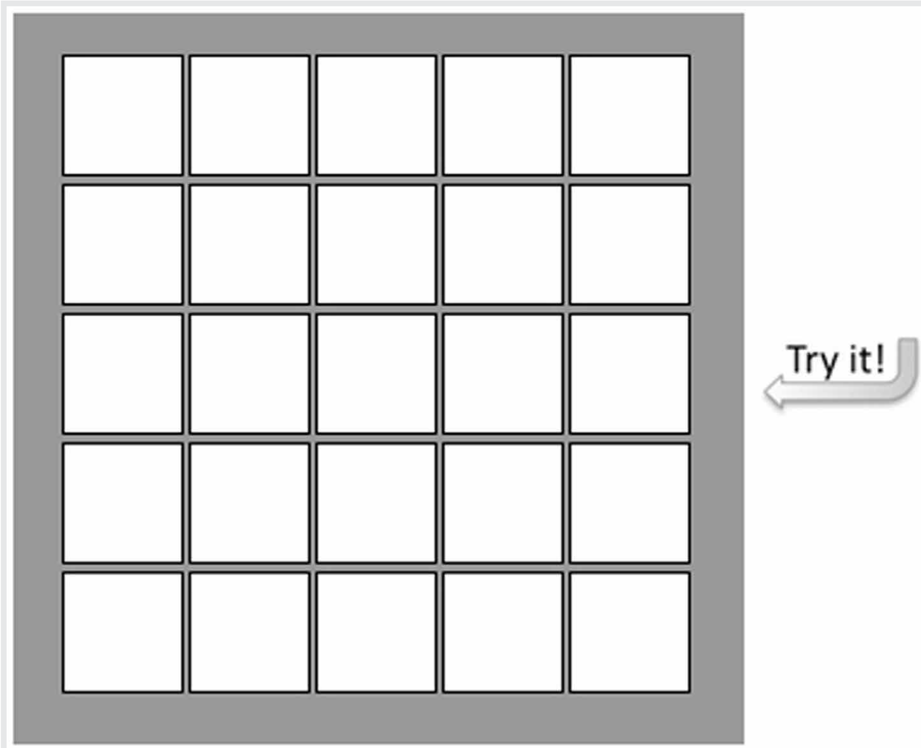



Figure 11.17: A 5x5 icon editor made out of individual pixels with toggle behaviour.

Real computers do not have two pixels, but millions. Figure 11.17 shows an example of how a matrix of elements with simple behaviour can become something potentially useful. Even with as few as 5x5 pixels you can use it to make basic shapes, i.e. an icon editor.

These kinds of explorations of interactive media can be extended in different directions. All the above examples have been with only two colours and discrete states. If we include analogue input, full colour, hidden states, delays, animations, sound, algorithms and communication through Internet, we get enough complexity to justify a whole new profession: Interaction Design.

## 11.8 IMPLICATIONS FOR INTERACTION DESIGNERS

What are the implications of this for interaction designers? We have gone through four perspectives on interactivity: interaction as information processing (Cognitive Science), interaction as tool use (Heidegger), interaction as perception (Merleau-Ponty), and interaction from a media and art perspective (Arnheim/Itten). The focus of the current analysis has mostly been on the interactivity of digital products. The implications for design will consequently mainly be related to the design of interactivity:

### 11.8.1 Interactivity is an important part of the user experience

- Combining the perspectives of Merleau-Ponty and Arnheim, it becomes evident that the interactivity of digital artefacts is perceived as *interaction gestalts* at a very immediate level, similar to that of visual perception. Interactivity is not simply the behaviour necessary to implement a certain functionality, but an important quality of a digital product. We often talk about the "look and feel" of a digital product. Users perceive "the feel" of the product through interaction, and this thus becomes an important part of the resulting user experience. Users care about "the feel".

### 11.8.2 A product's "feel" should be designed with care to detail

- If you want to design interactive products that stand out, you must be conscious of the interactive qualities of your products. "The feel" should consequently be designed; not only engineered. As with everything designed, God is in the details also concerning interactivity.

### 11.8.3 Perception of "the feel" requires action: Invite the user in

- Perception of the interactive dimension of a product, its "feel", requires user actions. It is therefore important that the product signals its potentials for interaction, its "affordances". The user experience is created through interaction. Interactivity that is hidden away is like a tree falling in the forest with no one watching.

### 11.8.4 Design for unforeseen use

- From a phenomenological perspective, objects get their meaning through use and social interaction. As Norman pointed out in his video on affordances, a good product should be designed for use situations other than the one intended.

### 11.8.5 Design for "the lived body"

- When designing for interaction beyond mouse and keyboard (e.g. mobile and whole-body interaction), design for the experienced body. This requires a focus on the bodily feeling of using the product.

### 11.8.6 Take responsibility for the feel of the total user experience

- In some cases, the user experience is a result of the user's interaction with a number of interconnected products and services, e.g. the combination of earplugs, an MP3 player, and an online music store. In those cases it is important that the "feel" of these products and services are designed in such a way that the sum give rise to a good user experience.

### 11.8.7 Interaction designers need to learn basic programming skills

- Designing interactivity requires the ability to make rapid "sketches" of interactive behaviour. This is important to be able to explore different behaviours, and to have running specification to hand over to the programmers. Despite numerous attempts to make the process of designing interactivity non-technical, interaction designers who want to add an extra quality to the interactive user experience still need to learn basic programming skills. Programming is the tool for shaping interactivity. A lot can be done with simple programming environments like Processing and Arduino.

### 11.8.8 The perceptual field is personal: Test with real users and listen carefully

- Make numerous sketches of interactive behaviour, but always test them on real users before you make important design decisions. Different users can perceive the same interactive behaviour in surprisingly different ways; and often very differently from you. This is not only because they interpret and experience their interactions differently, but also because their ways of interacting differ. Test for more than usability; ask them how it feels in use and listen carefully to what they tell you.

### 11.8.9 Intentionality and context matters: Make the tests realistic

- The perception of a product's interactivity is to a large extent coloured by the user's intention. Trying out "the feel" of a product in a controlled setting is very different from using it in a real context and for a real purpose. Ideally, tests should be done with real tasks and in real contexts. If that is not possible, you should be aware of the difference between real use and your test setting, and how this colours the user experience.

### 11.8.10 Interaction designers should be skilled in kinesthetic thinking

- The interactivity of digital products should be designed by interaction designers with a special sensitivity for "the feel" of a product.

> "The feel" is about action-reaction; it involves the whole body and it is about timing and rhythm. This requires the interaction designer to develop skills in *kinesthetic thinking* and *bodily intelligence*. While drawing classes are excellent for designing "the look", interaction designers should also consider classes in dance, drama or martial art to develop their sensitivity for things interactive.

Getting the feel right is of course not sufficient to make successful interactive products, but it is my belief that in a competitive market, products with a well designed feel will always stand out; interactivity matters.

## 11.9 WHERE TO LEARN MORE

### 11.9.1 Books

Winograd, Terry and Flores, Fernando (1986): *Understanding Computers and Cognition*. Norwood, NJ, Intellect

This book by Stanford professor Terry Winograd and former minister of finance of Chile, Fernando Flores, was primarily written as a criticism of artificial intelligence and cognitive science, but has strong relevance for a discussion of interactivity. As mentioned above, it presents three alternatives to cognitive science, of which the phenomenology of Heidegger is the most relevant here. Their interpretation of Heidegger was strongly influenced by Berkeley professor Hubert Dreyfus. The book was important by showing the relevance of continental philosophy to fundamental discussions in computer science.

Dourish, Paul (2001): *Where the Action Is: The Foundations of Embodied Interaction*. MIT Press

This book has become the standard textbook on the relevance of phenomenology for HCI. It starts out with a discussion of Ubiquitous Computing (pre-iPhone) and Social Media (pre-FaceBook), and shows that the phenonemology of Heidegger, Merleau-Ponty and Schutz is an excellent foundation for understanding the embodied and social nature of these technologies. He also draws on Wittgenstein and his concept of language games. Dourish introduces embodied interaction as a theoretical construct to capture the way in which interactive technologies become integral parts of our lives.

Dreyfus, Hubert L. (1990): *Being-in-the-World: A Commentary on Heidegger's Being and Time, Division I*. A Bradford Book

This is still one of the best introductions to the work of Martin Heidegger; from a philosopher who has been very successful in making continental philosophy available to Anglo-American readers. The book is a good starting point for a dive into 20th century German and French phenomenology.

Svanaes, Dag (2000): *Understanding Interactivity: Steps to a Phenomenology of Human-Computer Interaction*. Trondheim, Norway, Norges Teknisk-Naturvitenskapelige Universitet (NTNU)

Much of the current encyclopedia entry builds on my 2000 PhD monograph on interactivity. The book contains extended discussions of the four perspectives presented here, in addition to more background on Activity Theory, Etnomethodology, and Semiotics. It also presents an experiment on how users percieve abstract interaction, and some examples of ways to simplify the design of interactivity. A pdf version of the book can be downloaded here.

Fällman, Daniel (2003). *In Romance with the Materials of Mobile Interaction : A Phenomenological Approach to the Design of Mobile Information Technol-*

*ogy (PhD Thesis)*. Umeå Universitet http://daniel.fallman.org/publications.html#thesis

This thesis deals analytically and through design with the issue of Human-Computer Interaction (HCI) with mobile devices and mobile interaction. This subject matter is theoretically, methodologically, and empirically approached from two outlooks: a phenomenological and a design-oriented attitude to research. The book gives a very good overview of relevant aspects of the works of Husserl, Heidegger and Merleau-Ponty.

Hornecker, Eva (2004). *Tangible User Interfaces als kooperationsunterstützendes Medium*. Mathematik & Informatik, der Universität Bremen http://elib.suub.uni-bremen.de/diss/docs/E-Diss907_E.pdf

If you read German, this is a very interesting PhD on Tangible Computing strongly inspired by phenomenology. Hornecker is the author of the encyclopedia entry on Tangible Computing.

Ehn, Pelle (1988): *Work-Oriented Design of Computer Artifacts*. Stockholm, Arbetslivscentrum

Ehn's PhD from 1988 sums up his experience as one of the founders of the participatory design tradition in Scandinavia. His analysis builds on the theoretical frameworks of Marx, Heidegger and Wittgenstein. Heidegger is used to show the tool-like nature of software, and to argue for building tools to support skilled work.

## 11.9.2 Relevant papers

*Das, Anita, Faxvaag, Arild and Svanaes, Dag (2011): Interaction design for cancer patients: do we need to take into account the effects of illness and medication?. In: Proceedings of the 2011 annual conference on Human factors in computing systems 2011. pp. 21-24*

Hornecker, Eva (2005): A Design Theme for Tangible Interaction: Embodied Facilitation. In: Gellersen, Hans-Werner, Schmidt, Kjeld, Beaudouin-Lafon, Michel and Mackay, Wendy E. (eds.) Proceedings of the Ninth European Conference on Computer Supported Cooperative Work 18-22 September , 2005, Paris, France. pp. 23-43

Hummels, Caroline, Overbeeke, Kees and Klooster, Sietske (2007): Move to get moved: a search for methods, tools and knowledge to design for expressive and rich movement-based interaction. In Personal and Ubiquitous Computing, 11 (8) pp. 677-690

Larssen, Astrid Twenebowa, Robertson, Toni and Edwards, Jenny (2007): The feel dimension of technology interaction: exploring tangibles through movement and touch. In: Proceedings of the 1st International Conference on Tangible and Embedded Interaction 2007. pp. 271-278

Larssen, Astrid Twenebowa, Robertson, Toni and Edwards, Jenny (2006): How it feels, not just how it looks: when bodies interact with technology. In: Kjeldskov, Jesper and Paay, Jane (eds.) Proceedings of OZCHI06, the CHISIG Annual Conference on Human-Computer Interaction 2006. pp. 329-332

Larssen, Astrid Twenebowa, Robertson, Toni and Edwards, Jenny (2007): Experiential Bodily Knowing as a Design (Sens)-ability in Interaction Design. In: Desform - 3rd European Conference on Design and Semantics of Form and Movement Dec. 12-13, 2007, Newcastle upon Tyne, UK.

Lim, Youn-kyung, Lee, Sang-Su and Lee, Kwang-young (2009): Interactivity attributes: a new way of thinking and describing interactivity. In: Proceedings of ACM CHI 2009 Conference on Human Factors in Computing Systems 2009. pp. 105-108

Lim, Youn-kyung, Stolterman, Erik A., Jung, Heekyoung and Donaldson, Justin (2007): Interaction gestalt and the design of aesthetic interactions. In: Koskinen, Ilpo and Keinonen, Turkka (eds.) DPPI 2007 - Proceedings of the 2007 International Conference on Designing Pleasurable Products and Interfaces August 22-25, 2007, Helsinki, Finland. pp. 239-254

Lowgren, Jonas (2006): Articulating the use qualities of digital designs. In: Fishwick, Paul A. (ed.). "Aesthetic Computing (Leonardo Books)". The MIT Presspp. 383-403

Lowgren, Jonas (2009): Towards an articulation of interaction aesthetics. In New Review of Hypermedia and Multimedia, 15 (2) pp. 129-146

Moen, Jin (2005): Towards people based movement interaction and kinaesthetic interaction experiences. In: Bertelsen, Olav W., Bouvin, Niels Olof, Krogh, Peter Gall and Kyng, Morten (eds.) Proceedings of the 4th Decennial Conference on Critical Computing 2005 August 20-24, 2005, Aarhus, Denmark. pp. 121-124

*Ozenc, Fatih Kursat, Kim, Miso, Zimmerman, John, Oney, Stephen and Myers, Brad A. (2010): How to support designers in getting hold of the immaterial material of software. In: Proceedings of ACM CHI 2010 Conference on Human Factors in Computing Systems 2010. pp. 2513-2522*

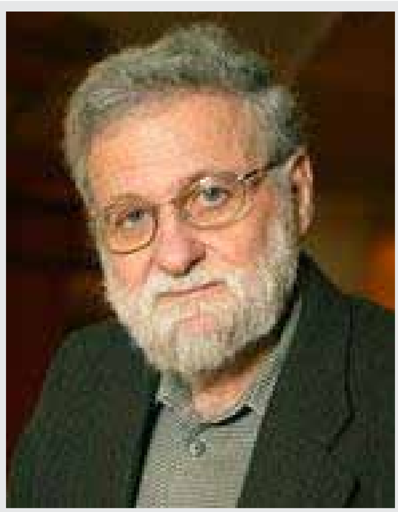
*Schiphorst, Thecla (2007): Really, really small: the palpability of the invisible. In: Proceedings of the 2007 Conference on Creativity and Cognition 2007, Washington DC, USA. pp. 7-16*

*Sundström, Petra and Höök, Kristina (2010): Hand in hand with the material: designing for suppleness. In: Proceedings of ACM CHI 2010 Conference on Human Factors in Computing Systems 2010. pp. 463-472*

## 11.12 COMMENTARY BY DONALD A. NORMAN

How to cite this commentary in your report

### Donald A. Norman



*© Donald A. Norman*

Don Norman is the author of numerous books including "Emotional Design," and more recently, "Living with Complexity." He is co-founder of the Nielsen Norman group, a professor at KAIST (in Korea), and IDEO fellow, and a design theorist, studying the fundamentals of modern design.Donald A. Norman has a Bachelor of Science in Electrical Engineering from MIT and a Doctor of Philosophy in Ps...

Donald A. Norman
Donald A. Norman is a member of The Interaction Design Foundation

It is clear that our understanding of the nature of interaction has been evolving in a practical sense, in terms of our scientific understanding, and in terms of the underlying philosophy. Dag Svanaes has done an excellent job of describing the changes in philosophical approaches to the study of interaction that have occurred in the last few decades. But the review feels as if it stopped at the end of the 20th century, but today we are well into the 21st century. So the changes described in the article have continued: we are still evolving our understanding, still making progress.

From my point of view, the study of interaction is far ranging, including all the ways by which we interact with the environment. As Svanaes points out, interaction means the interplay between the state of the environment and the actions of agents (which are usually people), where the actions change the state of the world. Interaction can be accidental or purposeful, goal-directed or deliberately non-intentional.

Although the definition provided by Svanaes is appropriately broad, the chapter and its examples are surprisingly limited to discussions of the way by which individuals interact with information systems. The chapter primarily reflects single individuals interacting with computational systems. Pity: interaction is far broader. Interaction designers must deal with groups of people, sometimes separated by time and or distance, with non-computational objects and with the natural world. Designers who are designing services, rules, and procedures must be concerned with how systems and services interact with employees, technical staff, and of course customers and others who make use of the services. Some interaction designers have to be concerned with interactions among machines. The principles of interaction design are far more important to be limited to single individuals interacting with information systems and displays.

Interaction design is critically important for lots of non-computer systems. Thus, skis are a tool for interaction, as are baseball and cricket bats, the rules of engagement of soccer (football), and the way by which we steer and control our automobiles. Interaction designers must work with a wide variety of technologies, not just those of information and computer systems.

Our philosophical basis for the understanding of interaction has been evolving rather dramatically. As Svanaes points out, in the early days of personal computers, which is where interaction design started, we took a strict, internal-processing, information processing point of view to describe the system and to develop rules for designers. This approach has been seminal in teaching us about affordances (and what I now call "signifiers"), conceptual models, and the role of feedback, appropriate mapping, and all the many other formal, theoretical principles that have resulted from these studies. However, this approach is also very narrow, limiting, and now widely considered either to be too restrictive or, by some, simply wrong.

Today, our understanding of human cognition is undergoing major changes. We no longer look at pure information processing. Instead, we know that emotion and cognition interact in fundamental ways, that the entire human body interacts with the environment, and that as a result we must come to understand the whole system. It is not enough to use limited input and output mechanism. Hence, the development of mobile systems that are location, position, and orientation sensitive, where the entire body is involved, where the gestures, posture, and movements of the entire body are taken as inputs to our systems and where the results are not simply changes in some image on a screen, but changes that envelop the body, surrounding it with sights, sounds, and feelings, exciting the entire panoply of sensory systems. Haptics and gestures are missing from this article as are simulators that move in space to engage the proprioceptive systems. Three-dimensional sound is as important as three-dimensional sight. Haptics is increasingly important.

The review of affordances in the article needs to be expanded. Affordances are about possibilities: an affordance is a relationship between a person and the environment. But it is not affordances that are critical to interaction, it is our perception of those affordances. After all, an affordance that is not known about might as well not exist. To make this distinction more clear, in my recent work I call the perceptual signal of an affordance a "signifier." (D. A. Norman, 2011: Living with Complexity. MIT Press.) Signifiers are the clues for people about possible actions, hence possible interactions.

I don't know where Svanaes got the notion that the popular view of information processing in HCI is that it "sees perception as sense data being passively received by the brain." (In his section on Merleau-Ponty). The notion that perception is an active, constructive process is an old one in psychological theories and it is certainly dominant today. We don't follow Merleau-Ponty: we follow where the data have taken us. But Gibson, for example, was a strong advocate of the active nature of exception, and that was 50 years ago.

The view of HCI and of cognitive science held by Svanaes seems rooted in the last century.

Today, we study interaction as a fundamental property of human and social behavior, rooted in the entire body, where the person – or people – is inside a sensory field that is rich in information. All the senses are involved, not just the visual sense that is the focus of this chapter: haptics, proprioception, smell, sight, and sound. That is why the development of rich sensors and novel haptics are so encouraging. This is why we have surround sound, motor and solenoid-driven hairs and other objects to grasp, fondle, and manipulate. This is why medical simulators involve the entire body, allowing the novice surgeon to put the hand inside the simulated body, feeling the beating heart, the throbbing veins, and the different pressures and feelings associated with cutting through different layers of skin. Interaction today is a rich, vibrant filed.

Some of these topics are still in their infancy, so the definitive articles within this Encyclopaedia will have to wait a few years. I am disappointed, however, that these new developments and approaches are not discussed within the article – they aren't even hinted at. So stay tuned for further news.

## 11.13 COMMENTARY BY EVA HORNECKER

How to cite this commentary in your report

### Eva Hornecker



Eva Hornecker is Assistant Professor at the University of Strathclyde. She has worked at several places before, including the UK's Open University, Sussex University, the Vienna University of Technology, following her PhD in Bremen, Germany. Eva's research focus is on 'Beyond the Desktop' Interaction Design. She researches in the intersections of UbiComp, tangible interfaces/interact...

Eva Hornecker
Eva Hornecker is a member of The Interaction Design Foundation

What is Interaction or Interactivity? Most HCI and Interaction Design textbooks do not really define (beyond more then one sentence) one of the most often used terms in our vocabulary. What seems so straightforward turns elusive when we try to dig deeper. We are in this situation probably because it is not evident where to turn to for definitions and theoretical underpinnings.

Different disciplines offer a large variety of definitions of *interaction*, ranging from phenomena of intersubjectivity and human sociality in social philosophy to system-theoretical approaches that analyze and model observable reciprocal action-effect relations and patterns. The word 'interaction' thus allows both for 'human as information processor' models, the kinds of formal state diagrams exemplified in this encyclopedia entry with a light switch, and 3d generation HCI approaches investigating the user experience of interaction. It even allows for anthropomorphic interpretations of human-computer interaction. Dag Svanaes provides a very good example of how we ascribe intentionality and agency to an animated object (the Mr. Peters button is evasive), even though the button itself merely acts on preprogrammed stimuli-response schemata. We perceive its behaviour – and our mind is basically desperately trying to see meaning in whatever we perceive and experience.

Svanaes' article traces the history of HCI's understanding of what happens on the user side of HCI. He steers the reader towards Phenomenology, starting with the Heideggerian view introduced by Winograd and Flores. The videos go into far more detail here – an important thought that seems missing from the written version is that ecologies of object derive meaning through each other.

The unique view that Svanaes contributes combines 'interaction as perception-action' (Merleau-Ponty) with elements of gestalt-perception, using very simple but powerful examples. He emphasizes how experiencing the 'feel' of interaction requires active engagement – the encyclopedia entry itself has interactive components that provide a first-hand experience. To experience this process, users must enter it; they need to actively interact in order to conceive the "dynamic gestalt" and the feel of interactive products. Feel is about dynamics, timing, and how the users' body is part of the interaction. One of the unusual recommendations that arises from this thought is that interaction designers should learn some form of sport or exercise that trains kinesthetic thinking and bodily awareness.

Svaneas beautifully explains (and illustrates) the part of interactivity that results from the perception-action loop – but it is hard to scale this up from his little colour-shifting disobedient buttons to large-scale behaving, intelligent, ambient objects and environments, with interactions that extend over long periods of time (and/or space). It is a bit like we smashed a mirror and now inspect a small crystal, which in miniature allows us to experiment with reflections, but that we find hard to put together so we can see the whole picture. I don't really have any recommendations on how to proceed from here – I think we are still only just beginning to understand what it means to have artificial interactive objects (that are not really living entities), and there may be further branches of philosophy to look at.

The view that Svanaes shares with many other contemporary interaction design protagonists is that the dynamics of interaction should be central in design. The experience of use is created through the interrelation of system behaviour and user activity. It is a dialogical and time-based, temporal process. I want to point out some of the literature in interaction design that I have found helpful for a better understanding of the longer-term process of interaction. These take a far more pragmatic and practical approach, without aiming to provide a philosophy of interaction, that may complement the approach taken in this article.

Crawford (2002) uses human conversation as a metaphor. It does not fully work, but turns out to be useful – the metaphor carried much further than I initially thought it was able to. It emphasizes how interactivity depends on having two active participants that contribute to the interaction and how the quality of interaction depends on what they are able to contribute, how well they understand each other, and what they can perceive of each other. The metaphor highlights that interaction can be skewed and dominated by one partner, or that one partner might be severely handicapped in their expressivity or ability to perceive and understand the other. Toni Robertson and her coworkers have done a similar analysis from a situated action perspective, analyzing what is available to the machine in an interaction process. Shedroff (2000) sees Interactivity as a continuum, between passive and interactive experiences. He points out aspects that make a system interactive, the most elemental components being system feedback and user control. Of more interest is that it often is the nature of the user activity that makes something interactive, e.g. in creativity tools or communication media. This means that a system that merely mediates messages between 2 points may be perceived as interactive, because it is an open system where the interactive component is 'donated' by the outside world, or, in the case of e.g. a drawing tool, by the user. Interaction transforms closed to open systems (cf. Wegner 1997) – even more so if interaction contributes content. Svanaes' article highlights how even within the seemingly mundane level where we only have user control and system feedback (the Mr. Peters button) interactional meaning emerges. Crawford and Shedroff motivate and might help us to analyse longer and more complex processes – does Mr Peters always react the same way, can we engage in a dialogue with him and what else can we do?

## References

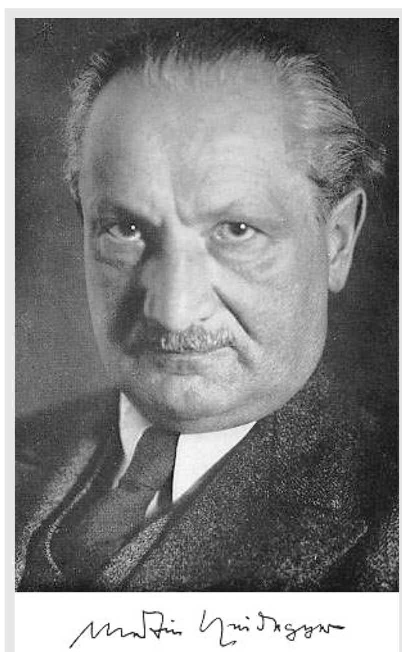

- Chris Crawford: The Art of Interactive Design: A Euphonious and Illuminating Guide to Building Successful Software. No Starch Press 2002
- Robertson, Toni and Loke, Lian. (2009) Designing Situations. Proceedings of OZCHI 2009, ACM. 1-8
- Nathan Shedroff: Information Interaction Design: A Unified Field Theory of Design. In: *Information Design,* Bob Jacobson (Hrsg.) MIT Press 2000
- Peter Wegner. Why Interaction is more powerful than Algorithms. CACM May 1997, Vol 40 Issue 5


## 11.10 RELATED IMAGES

Edmund Gustav Albrecht Husserl (April 8, 1859 - April 26, 1938) German philosopher and mathematician.

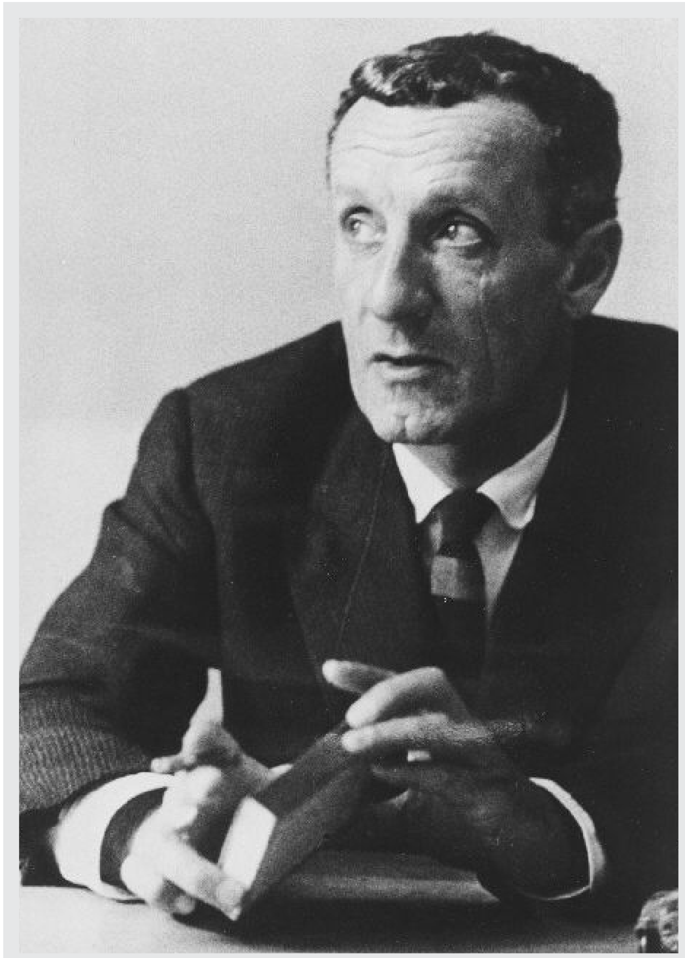
Ludwig Wittgenstein in his youth.





Heidegger.

*Courtesy of Taylor Carman.* 



Photograph of Maurice Merleau-Ponty.

Rene Descartes (31 March 1596 ? 11 February 1650): French philosopher.

## 11.11 BEHIND THE SCENES

*Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.*

*Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.*

*Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.*

*Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.*

*Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.*

## 11.14 REFERENCES


*Arnheim, Rudolf (1974): Art and Visual Perception: A Psychology of the Creative Eye. University of California Press*

*Card, Stuart K., Moran, Thomas P. and Newell, Allen (1983): The Psychology of Human-Computer Interaction. Hillsdale, NJ, Lawrence Erlbaum Associates*

*Elkoutbi, Mohammed and Keller, Rudolf K. (2000): User Interface Prototyping Based on UML Scenarios and High-Level Petri Nets. In: ICATPN 2000 2000. pp. 166-186*

*Harel, David (1987): Statecharts: A Visual Formalism for Complex Systems. In Sci. Comput. Program., 8 (3) pp. 231-274*

*Hartson, H. Rex and Gray, Philip D. (1992): Temporal Aspects of Tasks in the User Action Notation. In Human-Computer Interaction, 7 (1) pp. 1-45*

*Heidegger, Martin (1996): Being and Time (translated by Joan Stambaugh and Dennis J. Schmidt). State University of New York Press*

*Hornecker, Eva (2005): A Design Theme for Tangible Interaction: Embodied Facilitation. In: Gellersen, Hans-Werner, Schmidt, Kjeld, Beaudouin-Lafon, Michel and Mackay, Wendy E. (eds.) Proceedings of the Ninth European Conference on Computer Supported Cooperative Work 18-22 September , 2005, Paris, France. pp. 23-43*

*Hummels, Caroline, Overbeeke, Kees and Klooster, Sietske (2007): Move to get moved: a search for methods, tools and knowledge to design for expressive and rich movement-based interaction. In Personal and Ubiquitous Computing, 11 (8) pp. 677-690*

*ISO (2009). ISO 9241-210:2009. Ergonomics of human system interaction - Part 210: Human-centred design for interactive systems. International Organization for Standardization*

*Itten, Johannes (1974): The Art of Color: The Subjective Experience and Objective Rationale of Color. John Wiley and Sons*

*Larssen, Astrid Twenebowa, Robertson, Toni and Edwards, Jenny (2007): The feel dimension of technology interaction: exploring tangibles through movement and touch. In: Proceedings of the 1st International Conference on Tangible and Embedded Interaction 2007. pp. 271-278*

*Larssen, Astrid Twenebowa, Robertson, Toni and Edwards, Jenny (2007): Experiential Bodily Knowing as a Design (Sens)-ability in Interaction Design. In: Desform - 3rd European Conference on Design and Semantics of Form and Movement Dec. 12-13, 2007, Newcastle upon Tyne, UK.*

*Larssen, Astrid Twenebowa, Robertson, Toni and Edwards, Jenny (2006): How it feels, not just how it looks: when bodies interact with technology. In: Kjeldskov, Jesper and Paay, Jane (eds.) Proceedings of OZCHI06, the CHISIG Annual Conference on Human-Computer Interaction 2006. pp. 329-332*

*Lim, Youn-kyung, Stolterman, Erik A., Jung, Heekyoung and Donaldson, Justin (2007): Interaction gestalt and the design of aesthetic interactions. In: Koskinen, Ilpo and Keinonen, Turkka (eds.) DPPI 2007 - Proceedings of the 2007 International Conference on Designing Pleasurable Products and Interfaces August 22-25, 2007, Helsinki, Finland. pp. 239-254*

*Lim, Youn-kyung, Lee, Sang-Su and Lee, Kwang-young (2009): Interactivity attributes: a new way of thinking and describing interactivity. In: Proceedings of ACM CHI 2009 Conference on Human Factors in Computing Systems 2009. pp. 105-108*

*Loke, Lian, Larssen, Astrid Twenebowa, Robertson, Toni and Edwards, Jenny (2007): Understanding movement for interaction design: frameworks and approaches. In Personal and Ubiquitous Computing, 11 (8) pp. 691-701*

Lowgren, Jonas (2006): Articulating the use qualities of digital designs. In: Fishwick, Paul A. (ed.). "Aesthetic Computing (Leonardo Books)". The MIT Presspp. 383-403

Lowgren, Jonas (2009): Towards an articulation of interaction aesthetics. In New Review of Hypermedia and Multimedia, 15 (2) pp. 129-146

Lowgren, Jonas (2008). Interaction Design - brief intro. Retrieved 6 June 2013 from [URL to be defined - in press]

Merleau-Ponty, Maurice (1962): Phenomenology of Perception: An Introduction. Routledge

Moen, Jin (2005): Towards people based movement interaction and kinaesthetic interaction experiences. In: Bertelsen, Olav W., Bouvin, Niels Olof, Krogh, Peter Gall and Kyng, Morten (eds.) Proceedings of the 4th Decennial Conference on Critical Computing 2005 August 20-24, 2005, Aarhus, Denmark. pp. 121-124

Norman, Donald A. (1988): The Psychology of Everyday Things. New York, Basic Books

Ozenc, Fatih Kursat, Kim, Miso, Zimmerman, John, Oney, Stephen and Myers, Brad A. (2010): How to support designers in getting hold of the immaterial material of software. In: Proceedings of ACM CHI 2010 Conference on Human Factors in Computing Systems 2010. pp. 2513-2522

Schiphorst, Thecla (2007): Really, really small: the palpability of the invisible. In: Proceedings of the 2007 Conference on Creativity and Cognition 2007, Washington DC, USA. pp. 7-16

Stambaugh, Joan and Schmidt, Dennis J. (eds.) (1996): Being and Time (Suny Series in Contemporary Continental Philosophy). State Univ of New York Pr

Sundström, Petra and Höök, Kristina (2010): Hand in hand with the material: designing for suppleness. In: Proceedings of ACM CHI 2010 Conference on Human Factors in Computing Systems 2010. pp. 463-472

Svanaes, Dag (1993): Interaction is orthogonal to graphical form. In: Ashlund, Stacey, Mullet, Kevin, Henderson, Austin, Hollnagel, Erik and White, Ted N. (eds.) INTERACT 93 - IFIP TC13 International Conference on Human-Computer Interaction - jointly organised with ACM Conference on Human Aspects in Computing Systems CHI93 24-29 April, 1993, Amsterdam, The Netherlands. pp. 79-80

Svanaes, Dag (2000): Understanding Interactivity: Steps to a Phenomenology of Human-Computer Interaction. Trondheim, Norway, Norges Teknisk-Naturvitenskapelige Universitet (NTNU)

*Thimbleby, Harold (2004): User interface design with matrix algebra. In ACM Transactions on Computer-Human Interaction, 11 (2) pp. 181-236*

*Vogt, Stine and Magnussen, Svein (2007): Expertise in pictorial perception: eye-movement patterns and visual memory in artists and laymen. In Perception,*

*Winograd, Terry and Flores, Fernando (1986): Understanding Computers and Cognition. Norwood, NJ, Intellect*

## About the author

### Dag Svanaes

*© Dag Svanaes*

Has also published under the name of:
"Dag Svanaes"

Dag Svanaes is a professor at the Department of Computer and Information Science at the Norwegian University of Science and Technology. Svanaes is also adjunct professor at the IT-University of Copenhagen, Denmark He has been teaching and doing research in Human-Computer Interaction (HCI) since the late

1980s. His main areas of interest are mobile and ubiquitous computing, usability evaluation methodology, user-centered design, and the philosophy of interaction. He is currently involved in a national research initiative on medical informatics (NSEP), and have built up a usability lab for health ICT at NSEP.

## Your notes and thoughts on chapter 11

Record your notes and thoughts on this chapter. If you want to share these thoughts with others online, go to the bottom of the page at:
http://www.interaction-design.org/encyclopedia/philosophy_of_interaction.html

**NOTES:**

# Chapter 12

# Affective Computing

## Affective Computing, Affective Interaction and Technology as Experience

*by Kristina Höök.*

As Human-Computer Interaction (HCI) and Interaction Design moved from designing and evaluating work-oriented applications towards dealing with leisure-oriented applications, such as games, social computing, art, and tools for creativity, we have had to consider e.g. what constitutes an *experience*, how to deal with users' *emotions*, and understanding *aesthetic* practices and experiences. Here I will provide a short account of why in particular emotion became one such important strand of work in our field.

Video 12.1: Affective Computing video 1 - Introduction to Affective Computing and Affective Interaction.

*Courtesy of Rikke Friis Dam and Mads Soegaard. Copyright: CC-Att-ND (Creative Commons Attribution-NoDerivs 3.0 Unported). View full screen or download (736 KB). View transcription/captions.*

Video 12.2: Affective Computing video 2 - Main Guidelines and Future Directions.

*Courtesy of Rikke Friis Dam and Mads Soegaard. Copyright: CC-Att-ND (Creative Commons Attribution-NoDerivs 3.0 Unported). View full screen or download (448 KB). View transcription/captions.*

Video 12.3: Affective Computing video 3 - Designing Affective Interaction Products Dealing With Stress.

*Courtesy of Rikke Friis Dam and Mads Soegaard. Copyright: CC-Att-ND (Creative Commons Attribution-NoDerivs 3.0 Unported). View full screen or download (864 KB). View transcription/captions.*


Video 12.4: Affective Computing video 4 - Business value, value, and inspirations.

*Courtesy of Rikke Friis Dam and Mads Soegaard. Copyright: CC-Att-ND (Creative Commons Attribution-NoDerivs 3.0 Unported). View full screen or download (864 KB)*

I start by describing the wave of research in a number of different academic disciplines that resurrected emotion as a worthy topic of research. In fact, before then one of the few studies of emotion and emotion expression that did not consider emotion as a problem goes back as far as to Darwin's "The Expression of the Emotions in Man and Animals" in 1872 (Darwin, 1872). After Darwin, much attention in the academic world was focused on how emotion is problematic to rational thinking.

The new wave of research on emotion spurred ideas both amongst AI-researchers and HCI-researchers. In particular, the work by Rosalind Picard with her book on "Affective Computing" opened a viable research agenda for our field (Picard, 1997). But as with any movement within HCI, there will be different theoretical perspectives on the topic. A counter reaction to Picard's cognitivistic models of emotion came from the work by Sengers, Gaver, Dourish and myself (Boehner et al 2005,Boehner et al 2007, dePaula and Dourish 2005, Gaver 2009, Höök, 2008, Höök et al., 2008). Rather than pulling on a cognitivistic framework, this strand of work, Affective Interaction, draws upon phenomenology and sees emotion as constructed in interaction – between people and between people and machines.

While the work in these two strands on designing for emotion has contributed a lot of insights, novel applications, and better designs, both have lately come to a more realistic design aim where emotion is just one of the parameters we have to consider. Instead of placing emotion as the central topic in a design process, it is now seen as one component contributing to the overall design goal. In particular, it becomes a crucial consideration as we approach design for various experiences and interactions.

## 12.1 HISTORY: THE RESURRECTION OF EMOTION

During the 1990ies, there was a wave of new research on the role of emotion in diverse areas such as psychology (e.g. Ellsworth and Scherer, 2003), neurology (e.g. LeDoux, 1996), medicine (e.g. Damasio, 1995), and sociology (e.g. Katz, 1999). Prior to this new wave of research, emotions had, as I mentioned, been considered to be a low-status topic of research, and researchers had mainly focused on how emotion got in the way of our rational thinking. Results at that point focused on issues like when getting really scared, pilots would get tunnel vision and stop being able to notice important changes in the flight's surroundings. Being upset with a colleague and getting angry in the middle of a business meeting could sabotage the dialogue. Or giving a presentation and becoming very nervous could make you loose the thread of the argument. Overall emotions were seen the less valued pair in the dualistic pair rational – emotional, and associated with body and female in the "mind – body", "male – female" pairs. This dualistic conceptualisation goes back as far as to the Greek philosophers. In Western thinking, the division of mind and body was taken indisputable and, for example, Descartes looked for the gland that would connect the thoughts (inspired by God) with the actions of the body, Figure 12.1.

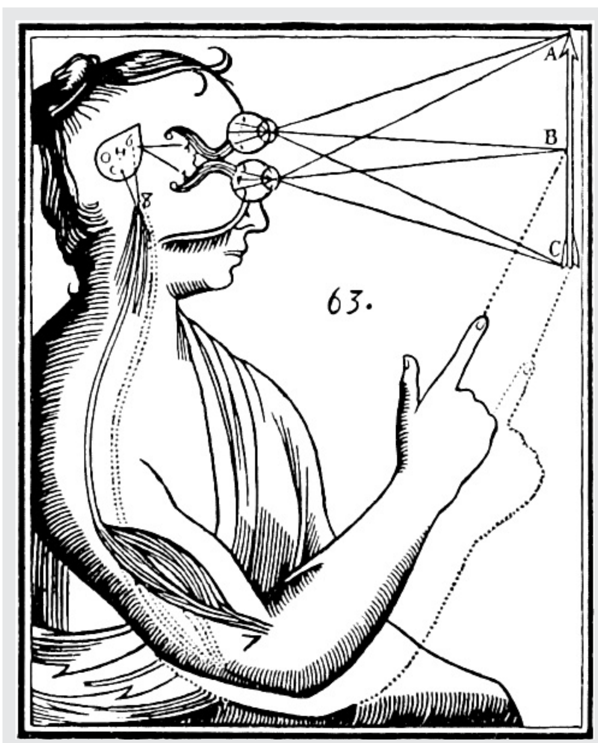



FIGURE 12.1: Ren? Descartes' illustration of dualism. Inputs are passed on by the sensory organs to the epiphysis in the brain and from there to the immaterial spirit.

But with this new wave of research in the 90ies, emotion was resurrected and given a new role. It became clear that emotions were the basis for behaving rationally. Without emotional processes we would not have survived. Being hunted by a predator (or enemy aircraft) requires focusing all our resources on escaping or attacking. Tunnel vision makes sense in that situation. Unless we can associate feelings of uneasiness with dangerous situations, as food we should not be eating, or people that aim to hurt us, we would make the same mistakes over and over, see Figure 12.2.

FIGURE 12.2: Focusing on enemy aircraft, getting tunnel vision.

*Courtesy of Master Sgt. Lance Cheung. Copyright: pd (Public Domain (information that is common property and contains no original authorship)).*

While fear and anger may seem as most important to our survival skills, our positive and more complex socially-oriented emotion experiences are also invaluable to our survival. If we do not understand the emotions of others in our group of primates, we cannot keep peace, share food, build alliances and friendships to share what the group can jointly create (Dunbar, 1997). To bring up our kids to function in this complex landscape of social relationships, experiences of shame, guilt, and embarrassment are used to foster their behaviour (Lutz 1986, Lutz 1988). But positive emotions also play an important role in bringing up our kids: conveying how proud we are of our kids, making them feel seen and needed by the adults, and unconditional love.

The new wave of research also questioned the old Cartesian dualistic division between mind and body. Emotional experiences are not residing in our minds or brains solely. They are experienced by our whole bodies: in hormone changes in our blood streams, nervous signals to muscles tensing or relaxing, blood rushing to different parts of the body, body postures, movements, facial expressions (Davidson et al., 2002). Our bodily reactions in turn feedback into our minds, creating experiences that regulate our thinking, in turn feeding back into our bodies. In fact, an emotional experience can start through body movements; for example, dancing wildly might make you happy. Neurologists have studied how the brain works and how emotion processes are a key part of cognition. Emotion processes are basically sitting in the middle of most processing going from frontal lobe processing in the brain, via brain stem to body and back (LeDoux, 1996), see Figure 12.3.

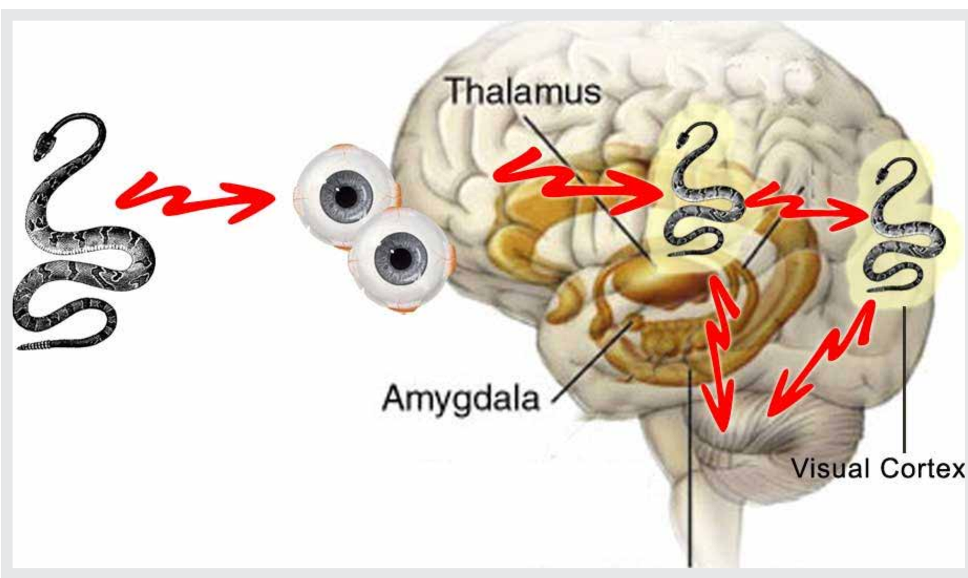



Figure 12.3: LeDoux's model of fear when seeing a snake.

Bodily movements and emotion processes are tightly coupled. As discussed by the philosopher Maxine Sheets-Johnstone in The Corporeal Turn: A interdisciplinary reader, there is "*a generative as well as expressive relationship between movement and emotion*" (Sheets-Johnstone, 2009). Certain movements will generate emotion processes and vice-versa.

But an emotional experience is not only residing "inside" our bodies as processes going back and forth between different parts of our body, they are also in a sense spread over the social setting we are in (Katz, 1999, Lutz, 1986, Lutz 1988, Parkinson, 1996). Emotions are not (only) hard-wired processes in our brains, but changeable and interesting regulating processes for our social selves. As such, they are constructed in dialogue between ourselves and the culture and social settings we live in. Emotion is a social and dynamic communication mechanism. We learn how and when certain emotions are appropriate, and we learn the appropriate expressions of emotions for different cultures, contexts, and situations. The way we make sense of emotions is a combination of the experiential processes in our bodies and how emotions arise and are expressed in specific situations in the world, in interaction with others, coloured by cultural practices that we have learnt. We are physically affected by the emotional experiences of others. Smiles are contagious.

Catherine Lutz, for example, shows how a particular form of anger, named *song* by the people on the south Pacific atoll Ifaluk, serves an important socializing role in their society (Lutz, 1986, Lutz 1988). Song is, according to Lutz, "justifiable anger" and is used with kids and with those who are subordinate to you, to teach them appropriate behaviour in e.g. doing your fair share of the communal meal, failing to pay respect to elders, or acting socially inappropriately.

Ethnographic work by Jack Katz (1999) provides us with a rich account of how people individually and group-wise actively produce emotion as part of their social practices. He discusses, for example, how joy and laughter amongst visitors to a funny mirror show is *produced* and regulated between the friends visiting together. Moving to a new mirror, tentatively chuckling at the reflection, glancing at your friend, who in turn might move closer, might in the end result in 'real' laughter when standing together in front of the mirror. Katz also places this production of emotion into a larger complex social and societal setting when he discusses anger among car drivers in Los Angeles, see Figure 12.4. He shows how anger is produced as a consequence of a loss of embodiment with the car, the road, and the general experience of travelling. He connects the social situation on the road; the lack of communicative possibilities between cars and their drivers; our prejudice of other people's driving skills related to their cultural background or ethnicity; etc.; and shows how all of it comes together explaining why anger is produced and when, for example, as we are cut off by another car. He even sees anger as a *graceful* way to regain a sense of embodiment.

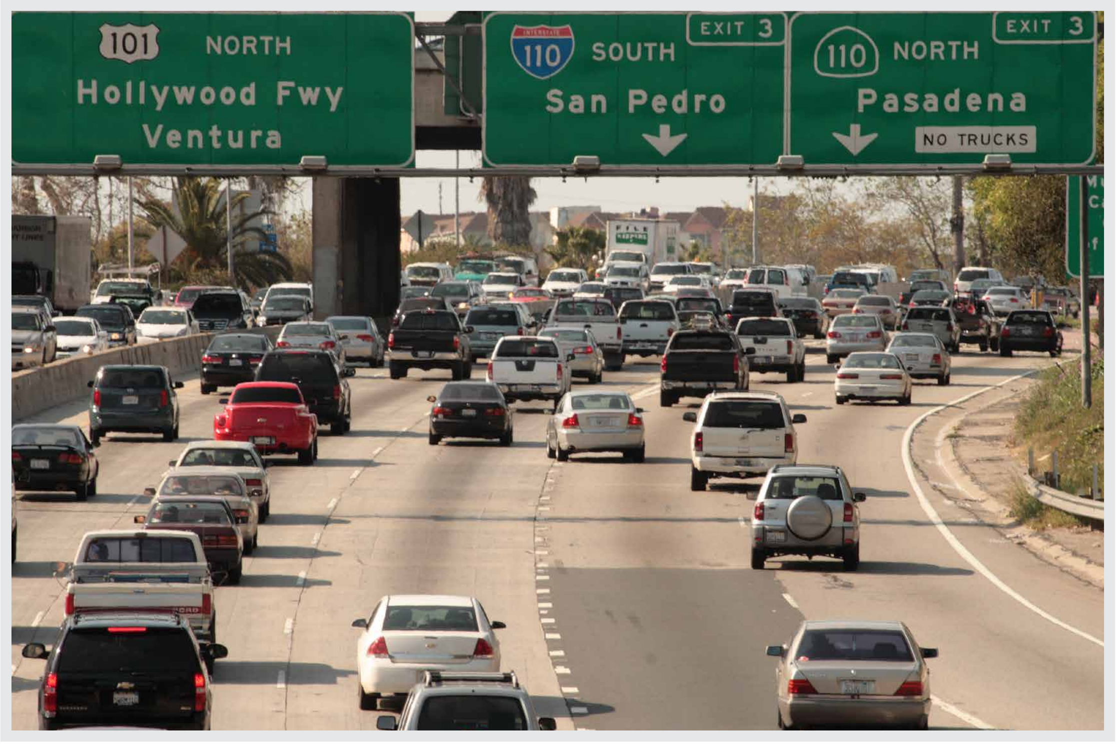



Figure 12.4: Katz places the production of emotion into a larger complex social and societal setting when he discusses anger among car drivers in Los Angeles.

## 12.2 EMOTION IN TECHNOLOGY?

A part of the new wave of research on emotion also affected research and innovation of new technology. In artificial intelligence, emotion had to be considered as an important regulatory process, determining behaviour in autonomous systems of various kinds, e.g. robots trying to survive in a dynamically changing world (see e.g. Cañamero, 2005).

In HCI, we understood the importance of considering users' emotions explicitly in our design and evaluation processes. Broadly, the HCI research came to go in three different directions with three very different theoretical perspectives on emotion and design.

1. The first, widely known and very influential perspective is that of Rosalind Picard and her group at MIT, later picked up by many other groups, in Europe most notably by the HUMAINE network. The cognitivistically inspired design approach she named *Affective Computing* in her groundbreaking book from 1997.

2. The second design approach might be seen as a counter-reaction to Affective Computing. Instead of starting from a more traditional perspective on cognition and biology, the *Affective Interaction* approach starts from a constructive, culturally-determined perspective on emotion. Its most well-known proponents are Phoebe Sengers, Paul Dourish, Bill Gaver and to some extent myself (Boehner et al., 2007, Boehner et al. 2005, Gaver 2009, Sundström et al. 2007, Höök, 2006, Höök 2008, Höök 2009).

3. Finally, there are those who think that singling out emotion from the overall interaction leads us astray. Instead, they see emotion as part of a larger whole of experiences we may design for – we can name the

movement *Technology as Experience*. In a sense, this is what traditional designers and artists have always worked with (see e.g. Dewey 1934) – creating for interesting experiences where some particular emotion is a cementing and congruous force that unites the different parts of the overall system of art piece and viewer/artist. Proponents of this direction are, for example, John McCarthy, Peter Wright, Don Norman and Bill Gaver (McCarthy and Wright, 2004, Norman, 2004, Gaver, 2009).

Let us develop these three directions in some more detail. They have obvious overlaps, and in particular, the Affective Interaction and Technology as Experience movements have many concepts and design aims in common. Still, if we simplify them and describe them as separate movements, it can help us to see the differences in their theoretical underpinnings.

### 12.2.1 Affective Computing

The artificial intelligence (AI) field picked up the idea that human rational thinking depends on emotional processing. Rosalind Picard's "Affective Computing" had a major effect on both the AI and HCI fields (Picard, 1997). Her idea, in short, was that it should be possible to create machines that relate to, arise from, or deliberately influence emotion or other affective phenomena. The roots of affective computing really came from neurology, medicine, and psychology. It implements a biologistic perspective on emotion processes in the brain, body, and interaction with others and with machines.

The most discussed and widespread approach in the design of affective computing applications is to construct an individual cognitive model of affect from what is often referred to as "first principles", that is, the system generates its affective states and corresponding expressions from a set of general principles rather than having a set of hardwired signal-emotion pairs. This model is combined with a model that attempts to recognize the user's emotional states through measuring the signs and signals we emit in face, body, voice, skin, or what we say related to the emotional processes going on. In Figure 12.5 we see for example how facial expressions, portraying different emotions, can be analysed and classified in terms of muscular movements.

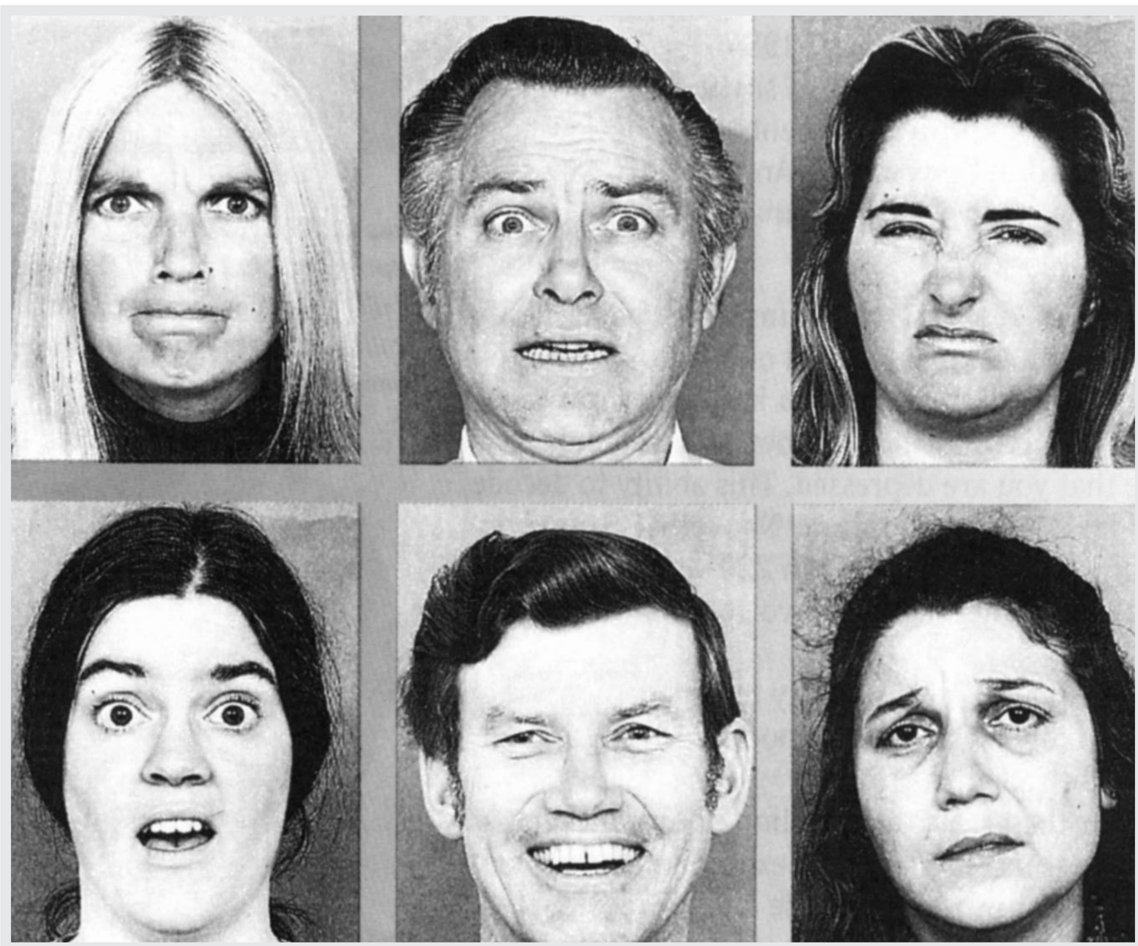


Figure 12.5: Facial expressions from Ekman portraying anger, fear, disgust, surprose, happiness and sadness.

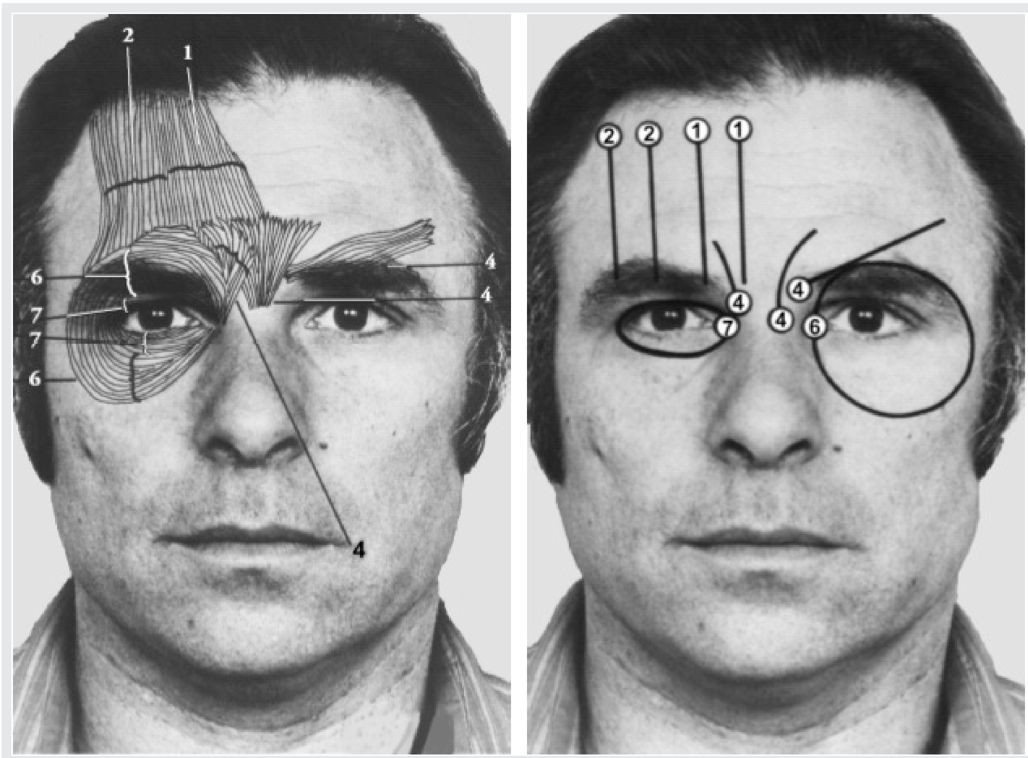


Figure 5B: Facial muscles moving eyebrow and muscles around the eye when expressing different emotions.



Emotions, or affects, in users are seen as identifiable states or at least identifiable processes. Based on the identified emotional state of the user, the aim is to achieve an interaction as life-like or human-like as possible, seamlessly adapting to the user's emotional state and influencing it through the use of various expressions. This can be done through applying rules such as those brought forth by Ortony et al. 1988, see Figure 12.6.

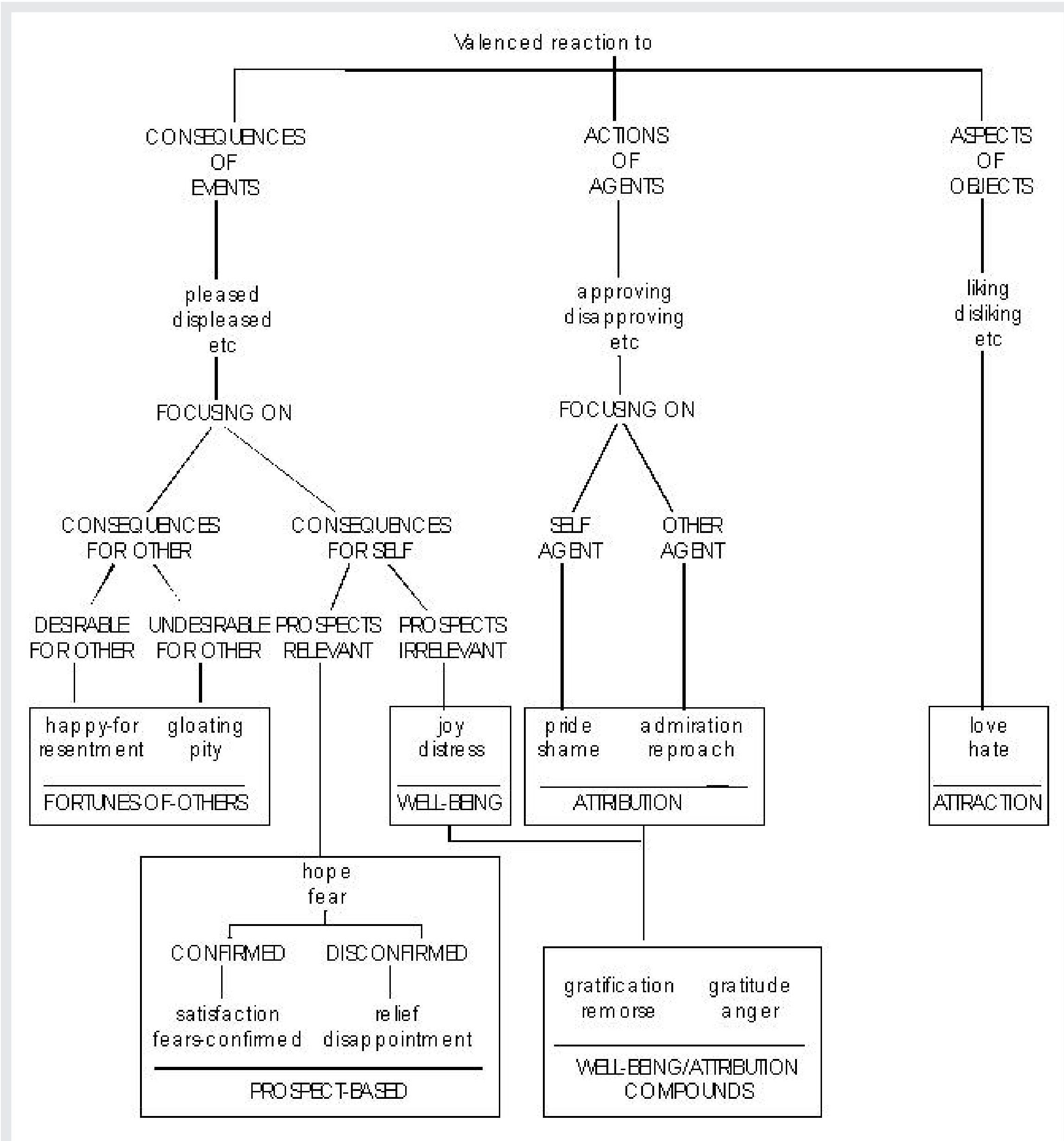



Figure 12.6: A rule from the OCC-model (Ortony et al., 1988).

This model has its limitations, both in its requirement for simplification of human emotion in order to model it, and in its difficult approach into how to infer the end-users emotional states through interpreting human behaviour through the signs and signals we emit. This said, it still provides for a very interesting way of exploring intelligence, both in machines and in people.

Examples of affective computing systems include, for example, Rosalind Picard and colleagues' work on affective learning. It is well known that students' results can be improved with the right encouragement and support (Kort et al., 2001). They therefore propose an emotion model built on James A. Russell's circumplex model of affect relating phases of learning to emotions, see Figure 12.7. The idea is to build a learning companion that keeps track of what emotional state the student is in and from that decides what help she needs.

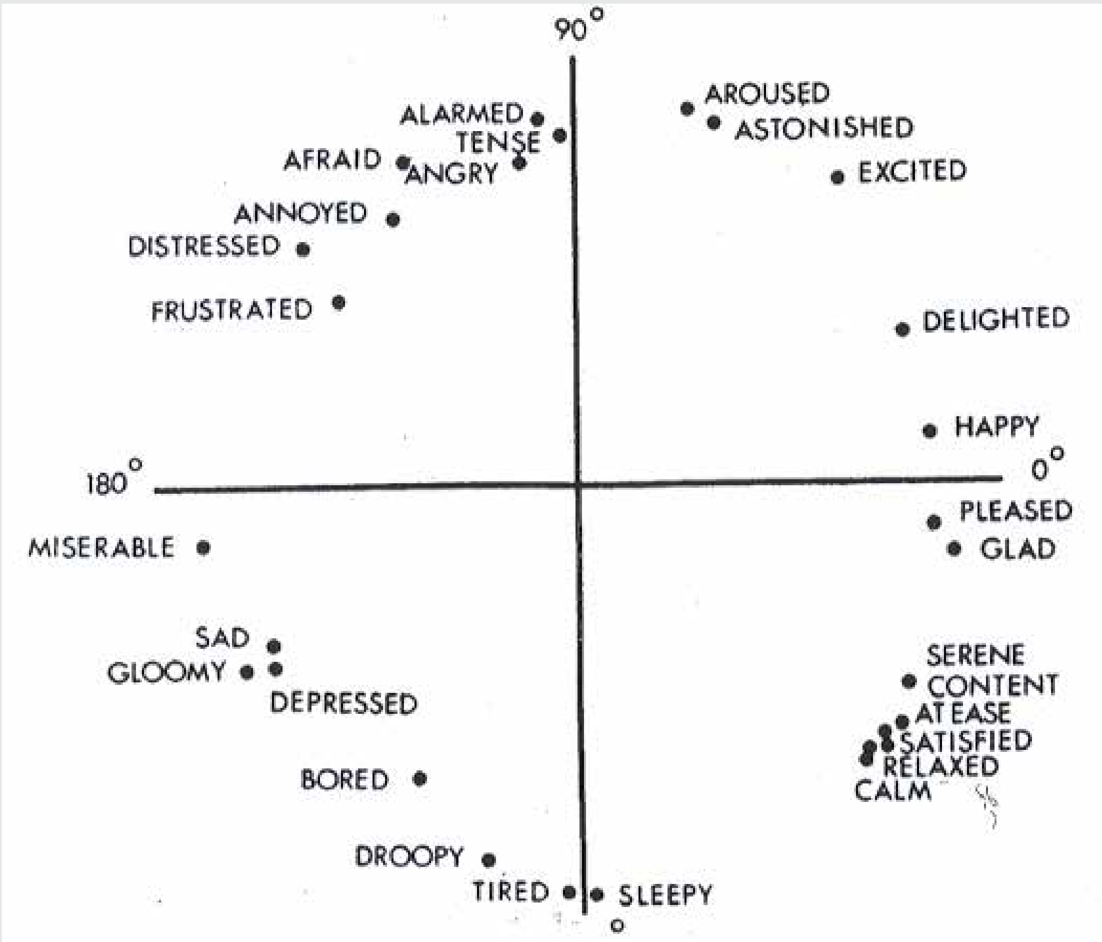



Figure 12.7: Russell's circumplex model of affect.



But the most interesting applications from Rosalind Picard's group deal with important issues such as how to train autistic children to recognise emotional states in others and in themselves and act accordingly. In a recent spin-off company, named Affectiva, they put their understanding into commercial use – both for the autistic children, but also for recognising interest in commercials or dealing with stress in call centres. A sensor bracelet recognising Galvanic Skin Response (GSR) is used in their various applications, see Figure 12.8.

FIGURE 12.8: The bracelet, named Q Sensor, measures skin conductance which in turn is related to emotional arousal - both positive and negative.



Other groups, like the HUMAINE network in Europe, starts from this way of seeing affective interaction.

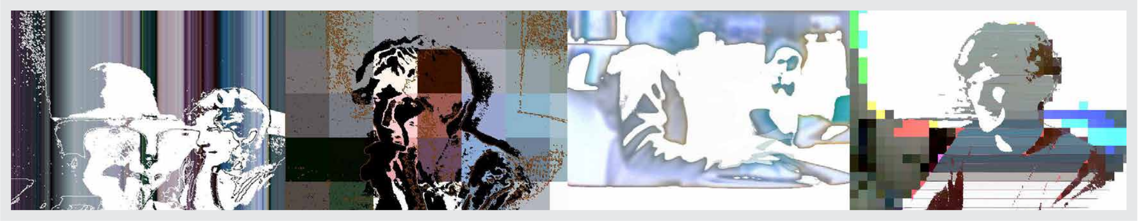


FIGURE 12.9: Samples of Affector Output.



### 12.2.2 Affective Interaction: The Interactional Approach

An affective interactional view is different from the affective computing approach in that it sees emotions as constructed in interaction, whereas a computer application supports people in understanding and experiencing their own emotions (Boehner et al., 2005, Boehner et al 2007, Höök et al., 2008, Höök 2008). An interactional perspective on design will not aim to detect a singular account of the "right" or "true" emotion of the user and tell them about it as in a prototypical affective computing application, but rather make emotional experiences available for reflection. Such a system creates a representation that incorporates people's everyday experiences that they can reflect on. Users' own, richer interpretation guarantees that it will be a more "true" account of what they are experiencing.

According to Kirsten Boehner and colleagues (2007)), the interactional approach to design:

1. recognizes affect as a social and cultural product
2. relies on and supports interpretive flexibility
3. avoids trying to formalize the unformalizable
4. supports an expanded range of communication acts
5. focuses on people using systems to experience and understand emotions
6. focuses on designing systems that stimulate reflection on and awareness of affect

Later, I and my colleagues added two minor modifications to this list (Höök et al., 2008):

1. Modification of #1: The interactional approach recognizes affect as an embodied social, bodily and cultural product
2. Modification of #3: The interactional approach is non-reductionist

The first change is related to the bodily aspects of emotional experiences. But explicitly pointing to them, we want to add some of the physical and bodily experiences that an interaction with an affective interactive system might entail.

We also took a slightly different stance towards design principle number three, "*the interactional approach avoids trying to formalize the unformalizable*", in Boehner and colleagues' list of principles. To avoid reductionist ways of accounting for subjective or aesthetic experiences, Boehner and colleagues' aim to protect these concepts by claiming that human experience is unique, interpretative, and ineffable. Such a position risks mystifying human experience, closing it off as ineffable and thereby enclosing it to be beyond study and discussion. While I wholeheartedly support the notion of unity of experience and support the idea of letting the magic of people's lives remain unscathed, I do believe that it is possible to find a middle ground where we can actually speak about qualities of experiences and knowledge on how to design for them without reducing them to something less than the original. This does not in any way mean that the experiential strands, or qualities, are universal and the same for everyone. Instead they are subjective and experienced in their own way by each user (McCarthy and Wright, 2004).

A range of systems has been built to illustrate this approach, such as Affector (Sengers et al., 2005), the VIO (Kaye, 2006), eMoto (Sundström et al., 2009), Affective Diary (Ståhl et al., 2009) and Affective Health (Ferreira et al., 2010) – just to mention a few.

Affector is a distorted video window connecting neighbouring offices of two friends (and colleagues), see Figure 12.9. A camera located under the video screen captures video as well as 'filter' information such as light levels, colour, and movement. This filter information distorts the captured images of the friends that are then projected in the window of the neighbouring office. The friends determine amongst themselves what information is used as a filter and various kinds of distortion in order to convey a sense of each other's mood.

eMoto is an extended SMS-service for the mobile phone that lets users send text messages between mobile phones, but in addition to text, the messages also have colourful and animated shapes in the background (see examples in

Figure 12.11). To choose an expression, you perform a set of gestures using the stylus pen (that comes with some mobile phones), which we had extended with sensors that could pick up on pressure and shaking movements. Users are not limited to any specific set of gestures but are free to adapt their gesturing style according to their personal preferences. The pressure and shaking movements can act as a basis for most emotional gestures people do, a basis that allows users to build their own gestures on top of these general characteristics, see Figure 12.11.

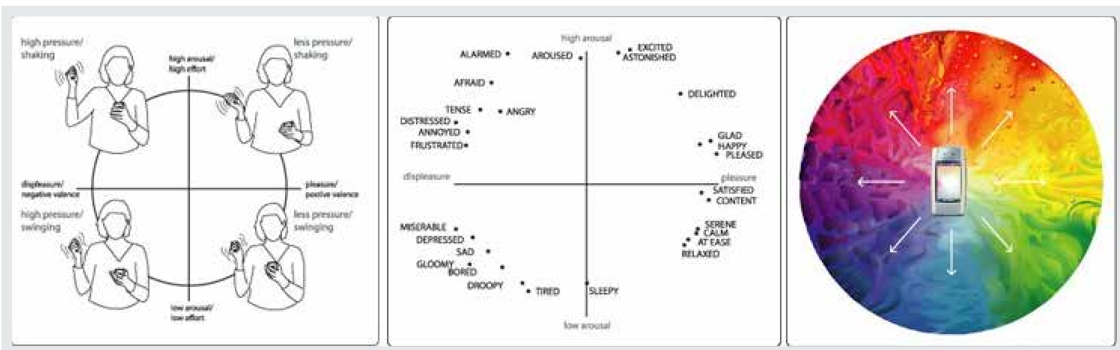



FIGURE 12.10: Different physical movements (left) that remind of the underlying affective experiences of the circumplex model of affect from Russell (middle), which is then mapped to a colourful, animated expression (right), also mapped to the circumplex model of affect.

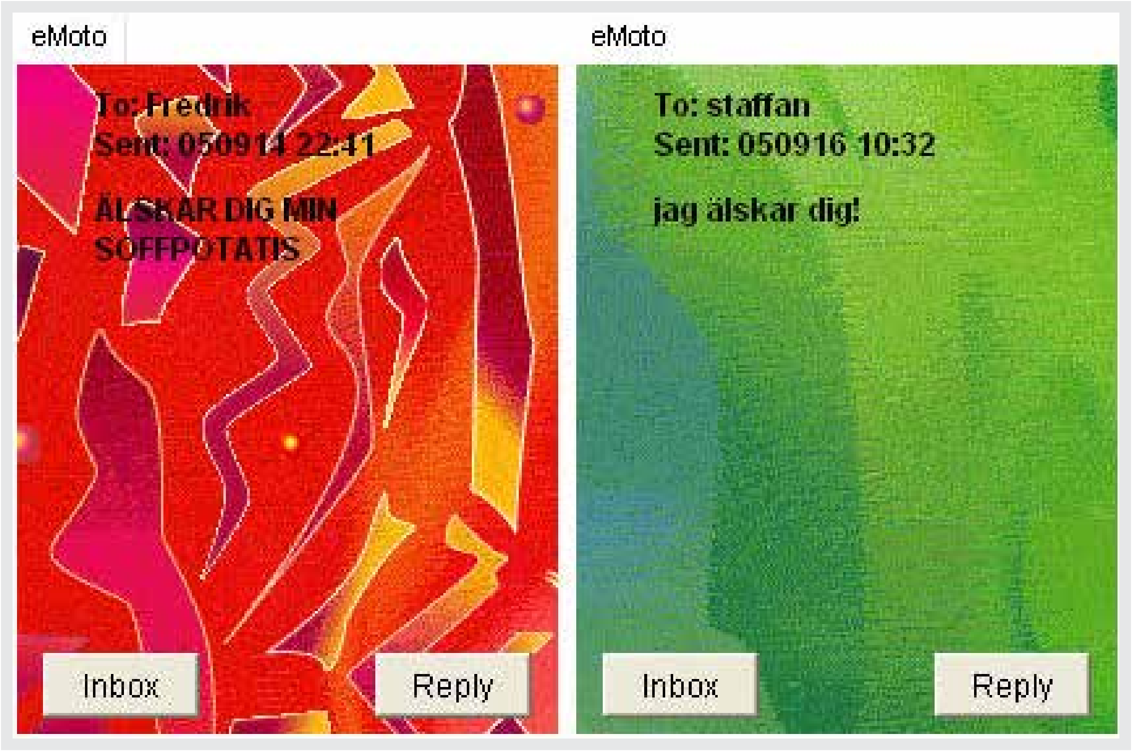



Figure 12.11: eMoto-messages sent to boyfriends in the final study of eMoto. On the left, a high energy expression of love from study participant Agnes to her boyfriend. On the right, Mona uses her favourite green colours to express her love for her boyfriend.



Affective Diary works as follows: as a person starts her day, she puts on a body sensor armband. During the day, the system collects time stamped sensor data picking up movement and arousal. At the same time, the system logs various activities on the mobile phone: text messages sent and received, photographs taken, and Bluetooth presence of other devices nearby. Once the person is back at home, she can transfer the logged data into her Affective Diary. The collected sensor data are presented as somewhat abstract, ambiguously shaped, and coloured characters placed along a timeline, see Figure 12.12. To help users reflect on their activities and physical reactions, the user can scribble diary-notes onto the diary or manipulate the photographs and other data, see example from one user in Figure 12.12.

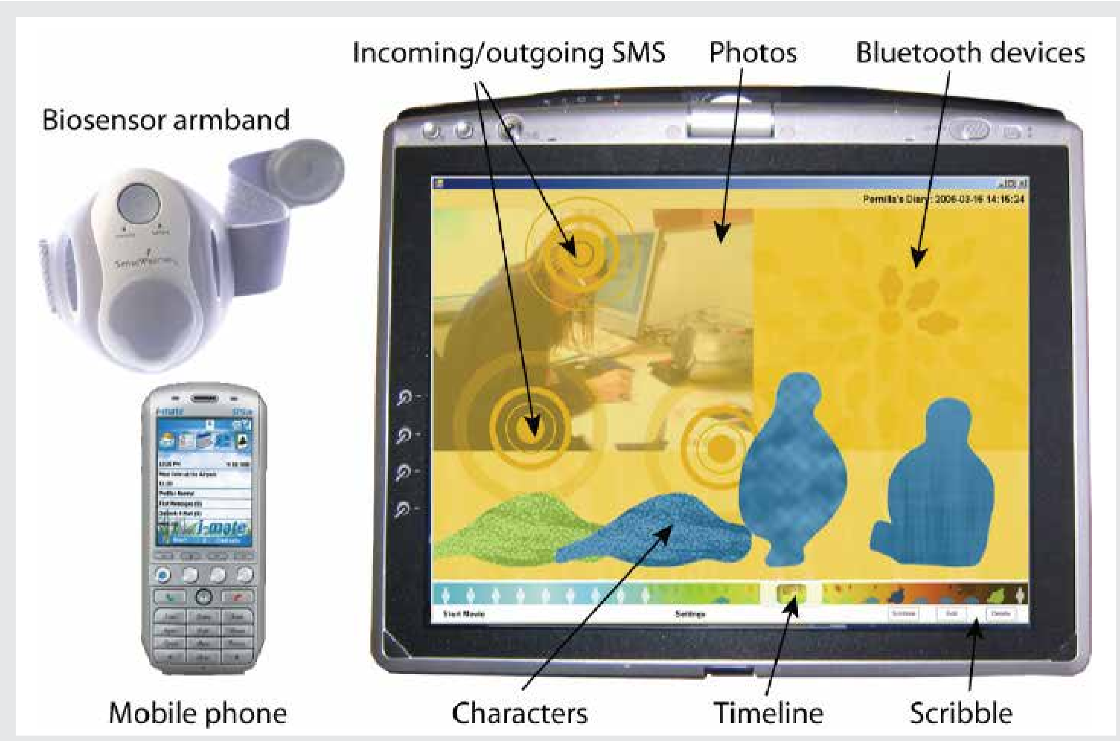



Figure 12.12: The Affective Diary system. Bio-sensor data are represented by the blobby figures at the bottom of the screen. Mobile data are inserted in the top half of the screen along the same time-line as the blobby characters.


Figure 12.13: A user says about this screendump: “[pointing at the orange character] And then I become like this, here I am kind of, I am kind of both happy and sad in some way and something like that. I like him and then it is so sad that we see each other so little. And then I cannot really show it.

*Copyright status: Unknown (pending investigation). See section “Exceptions” in the copyright terms.*

As can be seen from all these three examples, an interactional approach to design tries to avoid reducing human experience to a set of measurements or inferences made by the system to interpret users’ emotional states. While the interaction of the system should not be awkward, the actual experiences sought might not only be positive ones. eMoto may allow you to express negative feelings about others. Affector may communicate your negative mood. Affective Diary might make negative patterns in your own behaviour painfully visible to you. An interactional approach is interested in the full (infinite) range of human experience possible in the world.

### 12.2.3 Technology as Experience

While we have so far, in a sense, separated out emotion processes from other aspects of being in the world, there are those who posit that we need to take a holistic approach to understanding emotion. Emotion processes are part of our social ways of being in the world, they dye our dreams, hopes, and experiences of the world. If we aim to design for emotions, we need to place them in the larger picture of experiences, especially if we are going to address aspects of aesthetic experiences in our design processes (Gaver, 2009, McCarthy and Wright, 2004, Hassenzahl, 2008).

John Dewey, for example, distinguishes aesthetic experiences from other aspects of our life through placing it in-between two extremes on a scale (Dewey, 1934). On the one end of that scale, we just drift and experience an unorganized flow of events in everyday life, and on the other end of the scale we experience events that do have a clear beginning and end but that only mechanically connect the events with one-another. Aesthetic experiences exist between those extremes. They have a beginning and an end; they can be uniquely named afterwards, e.g. "*when I took those horseback riding lessons with Christian in Cambridge*" (Höök, 2010), but in addition, the experience has a unity – there is a single quality that pervades the entire experience (Dewey 1934, p. 36-57):

> "An experience has a unity that gives it its name, that meal, that storm, that rupture of a friendship. The existence of this unity is constituted by a single quality that pervades the entire experience in spite of the variation of its constituent parts."

In Dewey's perspective, emotion is (Dewey, 1934 p. 44):

> "the moving and cementing force. It selects what is congruous and dyes what is selected with its color, thereby giving qualitative unity to materials externally disparate and dissimilar. It thus provides unity in and through the varied parts of an experience."

However emotions are not static but change in time with the experience itself, just as a dramatic experience does (Dewey 1934, p. 43).

> "Joy, sorrow, hope, fear, anger, curiosity, are treated as if each in itself were a sort of entity that enters full-made upon the scene, an entity that may last a long time or a short time, but whose duration, whose growth and career, is irrelevant to its nature. In fact emotions are qualities, when they are significant, of a complex experience that moves and changes."

While an emotion process is not enough to create an aesthetic experience, emotions will be part of the experience and inseparable from the intellectual and bodily experiences. In such a holistic perspective, it will not make sense to talk about emotion processes as something separate from our embodied experience of being in the world.

Bill Gaver makes the same argument when discussing design for emotion (Gaver 2009). Rather than isolating emotion as if it is something that "*can be* *canned as a tomato in a Campbell tomato soup*" (as John Thackara phrased it when he criticised the work by Don Norman on the subject), we need to consider a broader view on interaction design, allowing for individual appropriation. Bill Gaver phrases it clearly when he writes:

> "Clearly, emotion is a crucial facet of experience. But saying that it is a 'facet of experience' suggests both that it is only one part of a more complex whole (the experience) and that it pertains to something beyond itself (an experience of something). It is that something—a chair, the home, the challenges of growing older—which is an appropriate object for design, and emotion is only one of many concerns that must be considered in addressing it. From this point of view, designing for emotion is like designing for blue: it makes a modifier a noun. Imagine being told to design something blue. Blue what? Whale? Sky? Suede shoes? The request seems nonsensical."

If we look back at the Affector, eMoto, and Affective Diary systems, we see clearly that they are designed for something else than the isolation of emotion. Affector and eMoto are designed for and used for communication between people where emotion is one aspect of their overall communication. And, in fact, Affector turned out to not really be about emotion communication, but instead became a channel for a sympathetic mutual awareness of your friend in the other office.

## 12.3 CONCLUDING REMARKS - SOME DIRECTIONS FOR THE FUTURE

It seems obvious that we cannot ignore the importance of emotion processes when designing for experiences. On the other hand, designing as if emotion is a state that can be identified in users taken out of context, will not lead to interesting applications in this area. Instead, the knowledge on emotion processing needs to be incorporated in our overall design processes.

The work in all the three directions of emotion design outlined above contributes in different ways to our understanding of how to increase our knowledge on how to make emotion processes an important part of our design processes. The Affective Computing field has given us a range of tools for both affective input, such as facial recognition tools, voice recognition, body posture recognition, bio-sensor models, and tools for affective output e.g. emotion expression for characters in the interface or regulating robot behaviours. The Affective Interaction strand has contributed to an understanding of the socio-cultural aspects of emotion, situating them in their context, making sure that they are not only described as bodily processes beyond our control. The Technology as Experience-field has shifted our focus from emotion as an isolated phenomenon towards seeing emotion processes as one of the (important) aspects to consider when designing tools for people.

There are still many unresolved issues in all these three directions. In my own view, we have not yet done enough to understand and address the everyday, physical, and bodily experiences of emotion processes (e.g. Sundström et al., 2007, Ståhl et al., 2009, Höök et al., 2008, Ferreira et al., 2008, Ferreira et al., 2010, Sundström et al., 2009, Ferreira and Höök, 2011). Already Charles Darwin made a strong coupling between emotion and bodily movement (Darwin, 1872). Since then, researchers in areas as diverse as neurology (leDoux 1996, Davidson et al., 2003), philosophy and dance (Sheets-Johnstone, 1999, Laban and Lawrence, 1974), and theatre (Boal, 1992), describe the close coupling between readiness to action, muscular activity, and the co-occurrence of emotion.

I view our actual corporeal bodies as key in being in the world, in creating for experiences, learning and knowing, as Sheets-Johnstone has discussed (1999). Our bodies are not instruments or objects through which we communicate information. Communication is embodied – it involves our whole selves. In design, we have had a very limited view on what the body can do for us. Partly this was because the technology was not yet there to involve more senses, movements and richer modalities. Now, given novel sensing and actuator materials, there are many different kinds of bodily experiences we can envision designing for – mindfulness, affective loops, excitement, slow inwards listening, flow, reflection, or immersion (see e.g. Moen, 2006, Isbister and Höök, 2009, Hummels et al., 2007). In the recently emerging field of design for somaesthetics (Schiphorst, 2007), interesting aspects of bodily learning processes, leading to stronger body awareness are picked up and explicitly used in design. This can be contrasted with the main bulk of e.g. commercial sports applications, such as pedometers or pulse meters, where the body is often seen as an instrument or object for the mind, passively receiving sign and signals, but not actively being part of producing them. Recently, Purpura and colleagues (2011) made use of a critical design method to pinpoint some of the problems that follows from this view. Through describing a fake system, Fit4Life, measuring every aspect of what you eat, they arrive at a system that may whisper into your ear "I'm sorry, Dave, you shouldn't eat that. Dave, you know I don't like it when you eat donuts" just as you are about to grab a donut. This fake system shows how we may easily cross the thin line from persuasion to coercion, creating for technological control of our behavior and bodies. In my view, by designing applications with an explicit focus on aesthetics, somaesthetics, and empathy with ourselves and others, we can move beyond impoverished interaction modalities and treating our bodies as mere machines that can be trimmed and controlled, towards richer, more meaningful interactions based on our human ways of physically inhabiting our world.

We are just at the beginnings of unravelling many novel design possibilities as we approach emotions and experiences more explicitly in our design processes. This is a rich field of study that I hope will attract many young designers, design researchers and HCI-experts.

## 12.5 COMMENTARY BY ROSALIND W. PICARD

How to cite this commentary in your report

### Rosalind W. Picard

© *Rosalind W. Picard*

Professor Rosalind W. Picard, Sc.D. is founder and director of the Affective Computing Research Group at the Massachusetts Institute of Technology (MIT) Media Laboratory, co-director of the Things That Think Consortium, the largest industrial sponsorship organization at the lab, and leader of the new and growing Autism & Communication Technology Initiative at MIT. She is co-founder, chie...

Rosalind W. Picard
Rosalind W. Picard is a member of The Interaction Design Foundation

This was an interesting chapter for me to try to understand and there is a banquet here for discussion, although I only have time to address one of the main dishes.

First, I want to say that I greatly appreciate the work of Kia Höök and others she cites to develop technologies for enhancing people's awareness of affect and helping people better reflect on and understand emotions, of self and others. I also deeply appreciate the work of designers to address holistic situations and design for people, including their feelings but also never *only* their feelings. These goals – creating interactions and designs that enhance affective understanding and that respond to the richness of human needs – are truly significant for improving much of what it means to be human. That said, I would like to correct an important misconception. Let me start with a story.

It was 1999 and Joe LeDoux, Antonio Damasio, and I had been invited to give talks on Emotion & Knowledge for the Barcelona Museum. The talks were simultaneously translated into multiple languages, giving me time to speak carefully and slowly in English, relying on the hard work of people more talented than I to translate into Catalan, Spanish, French, and more. It was a great experience overall – meeting fascinating people and engaging deeply in topics that were new and stimulating. But, there was one negative part that stands out in my memory. At the reception, a dark, trim, middle-aged man came striding in my direction, red-faced, furrowed brow, gesturing sharply, and having a hard time speaking. I've never seen somebody so angry in a museum. I glanced around me, thinking he was targeting somebody nearby who tried to steal his wife, after all, I was just nibbling on a canape. But his anger was at me. I swallowed, listened carefully, and gradually came to understand that in the language he was hearing my talk translated, he heard me claim something to the effect of "We have built or could now build human emotion into computers." I was actually extremely careful to NOT say that, but in his mind, I was denying the special feelings and experience we have that accompany human emotion, and reducing the great riches of our emotional experience entirely down to something like a text editor or game app. Listening to him, I realized that my careful choice of words in English, to say what we *were* doing precisely, and what I thought *could be done*, was translated inaccurately from my engineering culture, to his culture, which was social psychology.

In that reception, I learned, painfully, that what I meant by "modeling" was very different than what he heard when I said that word. I learned it was not enough to just be very careful with my words stating what we're doing. I needed to also anticipate how people from different fields could misinterpret what I said. I needed to learn to make additional clarifying remarks of what I did not mean. I should have said not only, "These are some of the *mechanisms related to* emotion that we are able to implement," but also, "These are *not* all of what emotion is." I should have said not only "by 'mechanisms of' I mean 'Attempts to represent'," but also, "Representing is not the same as reproducing." I did not realize he would otherwise be led to the wrong conclusion.

Why do I bring up this story? Höök's article refers to the Affective Computing approach as cognitivistic and reductionist, which is quite similar to the misunderstanding that happened in Barcelona in 1999.

When I speak or write of mechanisms of emotion, or models of emotion, I speak as an engineer trying to represent a complex phenomenon as best we can with tools we have: I do not confuse these representations with emotion itself. I am not a reductionist and Affective Computing is not reductionistic. I do not believe that emotion can be reduced to these representations, nor does Affective Computing claim this. I do not believe that emotion is "nothing but" the mechanisms we identify and build. The mechanisms we implement are not equivalent to the riches of human emotional experience, nor have I ever said that they will be: We have no evidence to make such claims. If people want to believe that emotions are entirely reducible to logical computation and bits, then that belief is based on faith, not science.

While people can *write about any concept using information and bits*, including emotion, I do not see evidence supporting the view that emotion can be fully reduced to bits and information. When I wrote Affective Computing, I knew many readers would be from AI, and would want to know how emotion might be implemented in machines, and so I described the parts of that process that I could envision. I was also very careful in my wording to not promote that such a method would be sufficient. However, I had not yet encountered the man in Barcelona, so one has to read my words carefully.

Unfortunately, if a person's views are multi-dimensional, people will try to reduce them to one dimension, and conveniently peg them on one hook or the other. The process is rather like tidying up the foyer by hanging each jacket on whatever hook happens to be available and strong enough to hold it up. Cognitivism is a handy hook, promoting the belief that thought can be fully reduced to rules and algorithms.

Cognitivism was a strong influence for AI pioneers like my friend and colleague, Marvin Minsky, who kept telling me "Emotions are just a special kind of thought", a sentence I disagreed with him on regularly and once get him to at least compromise by removing the word "just". Marvin believed bodies were irrelevant, except during infancy when people needed to be touched, else (studies showed) they withered and died. I have met other pioneers in AI who thought similarly. I am not of their camp, and my writings in Affective Computing talk about the body and about aspects of conscious experience that we haven't a clue how to implement in information and bits. There are some researchers who work in Affective Computing who hold a cognitivist view, but Affective Computing is not cognitivist.

Yes, Affective Computing includes some models and some researchers whose work might fit on a cognitivist hook, e.g. the cognitive rule-based models like OCC's could hang on a cognitivist hook for people who believe that approach could fully account for emotion (I don't). The stochastic signal-representing models of affect in speech or facial expression dynamics might hang on a different hook, and there are other hooks as well. The closet can be better organized than I have taken time to write about, especially as new garments keep arriving. But don't confuse the hooks with the house.

For some supportive examples, see “Chapter 1: Emotions are Physical and Cognitive” in Affective Computing (1997), containing some of my earliest writings on the need for a combined body-mind view in emotion research. That clearly does not fit on the cognitivist hook. Similarly, readers might be interested in the emphasis I placed on machines continusly co-creating interactions with people, taking into account not only emotion but also context and more, which resonates with the other areas Kia Höök’s article attempts to delineate (more examples are in “Chapter 8: Affective Wearables, see sections such as “Out of the lab and into the world.”) An affective technology does not have to use a formal AI model of emotion, or use discrete emotion recognition or a pattern classifier to fall under the area of affective computing.

But enough about organizing. I think the splitting and naming of pieces of a pie – whether it is an “affective computing pie” or some other kind of pie, is not as interesting as another question I see lurking behind the drive of some designers to separate themselves from a more objective engineering approach: Are emotions fully describable or are they ineffable?

In our work we have described emotion computationally and semantically, in numerous ways – discrete, dimensioned, numeric, semantic, as well as by quantifying creative behaviors, facial expressions, signal measurements of physiology and more. In no case do I think that we have “fully captured” human emotion with our models, methods, or descriptions. Something remains undescribed.

Affective computing often (but not always) tries to describe, objectively, more about emotion than has ever been described subjectively. Much of my work has pushed to make concrete, precise, in an engineering sense, measures of things that previously had only been addressed with words, self-report, questionnaires, whether applied to internal feelings or to outwardly observed behaviors. I am bothered by the way all subjective measurement methods are themselves influenced by emotion, and I want something more objective. Objective measures, however, do not imply reductionism any more than subjective measures imply reductionism. Both approaches “reduce” emotion to something – words, numbers, pictures, “blobby objects.” Using a representation is not being reductionistic. Reductionism is when people take an additional leap and say our emotions are “nothing but” what the computer is representing. The latter leap is one I have never promoted (except through mistranslated remarks).

When I closed my conversation with the man in Barcelona, we realized we were both deeply interested in better understanding emotion, and we realized that our perceived differences were actually not differences at all. Efforts to model do not imply a view of reductionism. Working to build representations that imitate some functions of emotions based on rules and categories does not mean cognitivism. Implementation of affective measures in bits does not mean emotion is only information. Affective computing creates tools toward greater goals – toward greater understanding of what makes us human. The man and I exchanged a hearty handshake and a smile before he departed.

I still have a lot to learn about communicating what we are trying to do with emotion – it’s a big topic, and it’s not one that just an engineering approach can conquer. I’m thrilled, as an engineer, to be sharing the journey with people from social psychology, design, neuroscience, AI, as well as many other arts and sciences. Together we’ll figure out much more than if we set up different camps.

The original definition I gave of affective computing is broader than the one Kia paraphrases: *Computing* (includes machines, robots, phones, sensors, smart clothing, anything that can do computation) that relates to, arises from, or deliberately influences emotion or other affective phenomena. This was never just about AI or HCI, or about making intelligent machines, although those were the largest communities I was trying to convince to work on emotion at the time.

Perhaps I can be permitted to close, using the opening I wrote in 1997, which still rings true today:

" In the course of this work I have come to appreciate all the more our own human needs for emotional development. It is my hope that this direction of research will encourage and enable us in this development – by no longer ignoring human emotions in human-computer interaction, by helping us become more aware of how we communicate, by providing testbeds for theories of emotion in learning and other functions, through animation of emotional characters and playful scenarios with which children can interact, by assisting scientists in collecting affective patterns, by helping advance research on understanding the role of emotion in preventive medicine, and more. It is my hope that affective computers, as tools to help us, will not just be more intelligent machines, but will also be companions in our endeavors to better understand how we are made, and so enhance our own humanity. (Preface to Affective Computing, 1997) "

## 12.6 COMMENTARY BY PAUL HEKKERT

How to cite this commentary in your report

### Paul Hekkert

© *Paul Hekkert*

Paul Hekkert is professor of Form Theory at the department of Industrial Design of Delft University of Technology. His main research interest is product experience, including product aesthetics, emotion, expressiveness, and attachment. Next, he is involved in design methodology and has co-developed an interaction-centred design approach, called ViP (Vision in Product design)....

Paul Hekkert
Paul Hekkert is a member of The Interaction Design Foundation

Affective computing is an exiting discipline and Kristina Höök offers us some nice examples of what the field can bring. Wouldn't it be great if intelligent machines could somehow 'sense' what we feel when interacting with them and then adjust their actions accordingly? This is actually what the pioneers of affective computing saw as their challenge and this is what they have been after:

1. Is it possible to recognize people's emotional responses/states from their behavior, physiological responses or facial expressions, and

2. Can we make the system (e.g. computer, product, mobile device) take this information into account in appropriate responses? Rosalind Picard will correct me if I am wrong that these were and still are the main challenges of the AC discipline.

Is this reductionist? Sure it is, you can only measure a few indicators of people's emotional responses and each and every indicator (e.g. pressure exerted, skin conductance, heart rate variability, facial muscles) only tells a little part of the story. There are many behavioral, physiological and psychological sides to an emotion and we simply cannot tap them all. But what is the alternative? We want our measurements to be as non-invasive as possible. If we end up affecting people's behavior or even change their emotional states because of the way we are measuring, the whole purpose is lost. Preferably, we measure user's emotional responses without them being aware of it. The question is what each and every indicator of an emotion actually tells us (its validity) and how accurately – and unobtrusively – we can measure it. A lot of the work in AC has been put into these questions.

As an alternative, Kristina Höök proposes the "Interactional Approach" where the system allows users to reflect on their emotional experiences. This, however, does not eliminate the measurement problem as we can for example see in the "Affective Diary". This application registers movement and arousal as indicators of people's emotional state and transfers these data into shapes and colors as a form of feedback. If you aim to give people feedback on their feelings this is the 'appropriate response' you decided on and you have moved to the second challenge of AC: how to respond? And of course, the response you aim for very much depends on the type and function of the system. When I am typing a document in Word – as I am doing now – I do not want the system to give me continuous feedback on my emotional states, nor do I want to see these reflected in the words

I type. But I may want the system to recognize that I am in a hurry, or impatient, or stressed, and subtly 'make me' slow down, without me being aware of it. Miguel Bruns Alonso recently explored this idea in the design of a pen that senses implicit behaviors related to restlessness and responds by providing inherent feedback to lower the stress level (Bruns Alonso et al., 2011).

So, is there an alternative to the measurement problem? In another example from Kristina Höök's chapter, she describes eMoto, an extended SMS-service that allows people to communicate their feelings in colorful and animated shapes. This type of 'measurement' – making users express their own emotions in words or images – only works when we are dealing with communication devices and is problematic for different reasons. First of all, it *is* obtrusive and not very recommendable if communicating emotions is not the design goal. But also, there are validity problems in people's verbal or non-verbal reports of their own emotions. All kind of social rules, demand characteristics (of the device?), and response styles may interfere with a valid report of your own feelings (see e.g. Mauss and Robinson, 2009 for a review of emotion measures).

And yes, I fully agree with Kristina that the role of our body is relatively unexplored in the AC field and offers a lot of potential, both to the recognition and response challenge. Given recent advances in cognitive science (see e.g. Johnson, 2007), where bodily experiences are increasingly recognized as being at the roots of our thinking and feeling, we may expect more and more studies – like Bruns Alonso's – in which our body is the main mediator. How else can we "grasp" the affective domain?

## References


- Bruns Alonso, M., Hummels, C.C.M., Keyson, D.V., & Hekkert, P. (Conditionally accepted). Measuring and adapting behavior during product interaction to influence affect. *Personal and Ubiquitous Computing*.

- Johnson, M. (2007). *The meaning of the body: Aesthetics of human understanding*. Chicago: The University of Chicago Press.
- Mauss, I.B. & Robinson, M.D. (2009). Measures of emotion: A review. *Cognition and Emotion*, *23*, 209-237.


## 12.7 COMMENTARY BY EGON L. VAN DEN BROEK

How to cite this commentary in your report

### Egon L. van den Broek

*© Egon L. van den Broek*

Egon L. van den Broek, MSc (2001) in artificial intelligence, PhD (2005) in image retrieval, and PhD (2011) in affective computing. He is consultant and assistant professor (University of Twente and Radboud University Medical Center Nijmegen, The Netherlands). He guided 50+ students, has 150+ scientific publications, 5 patent applications, and various awards. Egon serves in boards of a...

Egon L. van den Broek
Egon L. van den Broek is a member of The Interaction Design Foundation

## On the bodily expressions of emotion, be aware: More than a century of research!

When thinking about this commentary, ideas popped up and emotions emerged. What to comment on? Kia Höök delivered an excellent chapter. She mentions three angles to approach emotion in technology from (cf. Van den Broek, 2011), namely: affective computing, affective interaction, and technology as experience. In this commentary, I will narrow the focus to affective computing solely. Furthermore, I have also chosen to take a step back and be so bold as to take a methodological perspective with a historical flavor. Why? Well, throughout the years I have discovered more and more literature that touches the core of affective computing but appears to be unknown (e.g., Arnold, 1968; Candland, 1962; Dunbar, 1954). This commentary is founded on two books from a time long before the term affective computing was coined, the 50s and 60s of the previous century. Both books are taken from completely distinct branches of science. Knowledge on science's history can prevent us, both practitioners and scientists, from repeating mistakes. As such, this commentary touches upon the essence of science itself.

Kia Höök provides a concise overview of emotion in technology. She embraced affective interaction instead of affective computing. In contrast, in this commentary, I have taken the affective computing standpoint. Moreover, Kia Höök has taken a design perspective, where this commentary touches upon and questions the fundaments of emotions in technology. Lessons had been learned but have already been forgotten (Arnold, 1968; Candland, 1962; Dunbar, 1954). Consequently, affective computing tends to reinvent the wheel, at least to some extent. Yes, this is a bold claim, a very bold claim but I hope that after reading this commentary, you as a reader may share my concerns.

In 1954, 5 years before her death, Flanders Dunbar delivered the fourth edition of "Emotions and bodily changes: A survey of literature on psychosomatic interrelationships 1910-1953". With this impressive volume, she provides an exhaustive and structured review of scientific literature of (roughly) the first half of the previous century on emotions and bodily changes. The volume's title is well chosen and reflects its content nicely. This makes this book undoubtedly valuable for the community of affective computing. However, as far as I know, outside my own work (e.g., Van den Broek, 2011), not a single reference is made to this book in any affective computing article, report, or book. I can only hope that I have missed quite a few …

Flanders Dunbar starts her book (1954) with:

> "Nearly half a millennium B.C., Socrates came back from army service to report to his Greek countrymen that in one respect the barbarian Thracians were in advance of Greek civilization: They knew that the body could not be cured without the mind. "This," he continued, "is the reason why the cure of many diseases is unknown to the physicians of Hellas, because they are ignorant of the whole." It was Hippocrates, the Father of Medicine, who said: "In order to cure the human body it is necessary to have knowledge of the whole of things." And Paracelsus wrote: "True medicine only arises from the creative knowledge of the last and deepest powers of the whole universe; only he who grasps the innermost nature of man, can cure him in earnest." To us today this seems rather an impossible demand (p. 3)."

Where the work of Dunbar illustrates that the origins of affective computing can be traced back to more than a century ago, this quote illustrates that knowledge on the interaction between body and mind was already known more than 25 centuries ago! Let us now identify some core concepts as mentioned in the quote from Dunbar (1954), which are crucial for affective computing.

The old Greek already noted that "*the body could not be cured without the mind*" (cf. Kia Höök's chapter). So, both are indisputably related and, hence, in principle, the measurement of emotions should be feasible. This is well illustrated by the remark that "*the cure of many diseases is unknown to the physicians of Hellas*", as the Greek culture was devoted to the body and not to the mind. Recent work confirmed this relation. For example, when chronic stress is experienced, similar physiological responses emerge as were present during the stressful events from which the stress originates. If such physiological responses persist, they can cause pervasive and structural chemical imbalances in people's physiological systems, including their autonomic and central nervous system, their neuroendocrine system, their immune system, and even in their brain (Brosschot, 2010). This brings us to the need for the "*knowledge of the whole of things*", a holistic view, perhaps closely related to what Kia Höök denotes as Technology as Experience. Although the previous enumeration of people's physiological systems can give the impression that we are close to a holistic model, it should be noted that this is in sharp contrast with the current level of science. For example, with (chronic) stress, a thorough understanding is still missing. This can be explained by the complexity of human's physiological systems, the continuous interaction of all systems, and their integral dynamic nature. However, Brosschot (2010) considers emotions as if these can be isolated and attributed to bodily processes only. I firmly agree with Kia Höök that dynamics beyond the body should also be taken into account. Moreover, as Kia Höök also notes, in relation to computing entities, the interaction consists of much more than emotions; however, the same is true when no computing is involved at all.

25 centuries ago scientists did not apply modern statistics; however, 1 century ago, scientists did already apply statistics; for example, Fisher invented the ANOVA class of statistical models in 1918. This provided the means to test and

generalize findings on emotions and bodily changes and boosted the development of behavioral sciences in general (Dunbar, 1954). Moreover, this work fits into Rosalind W. Picard's definition of affective computing: "*... a set of ideas on what I call "affective computing," computing that relates to, arises from, or influences emotions.*" (Picard, 1995, p. 1) At least it fits when taken as the traditional interpretation of computing (i.e., to determine by mathematical means). However, the added value of affective computing would be its engineering component, in particular, signal processing and pattern recognition (Van den Broek, 2011). This would enable machines to sense emotions, reason about them, and perhaps develop them themselves. This would mark a new era of computing.

With the invention of computing machinery, shortly after World War II, a new type of statistics was developed: pattern recognition. In his edited volume "Methodologies of Pattern Recognition" (1969), Satosi Watanabe collected a set of papers that were presented or meant to be presented at the International Conference on Methodologies of Pattern Recognition in 1968. Watanabe started his book with defining pattern recognition:

> " To the layman's ear, the term pattern recognition sounds like a very narrow esoteric field of electronic computer applications. But, actually, it is a vast and explicit endeavor at mechanization of the most fundamental human function of perception and concept formation (p. vii). "

Watanabe denotes pattern recognition by computers as the "*mechanization of the most fundamental human function of (i) perception and (ii) concept formation.*" Up to this date human pattern recognition in general is largely unsolved. We do not understand how we, as humans, process affective signals (Van den Broek, 2011). Moreover, the perception of signals and, subsequently, patterns is one thing; their interpretation in terms of emotions is something completely different. This issue refers to content validity; that is, (i) the agreement among experts on the domain of emotions; (ii) the degree to which a (low level) percept adequately represents an emotion; and (iii) the degree to which (a set of) percepts adequately represents all aspects of the emotions under investigation.

The issue of concept formation relates to the process of construct validation, which aims to develop a ground truth (or an ontology or semantic network), constructed around the emotions investigated. Such a framework requires theoretically grounded, observable, operational definitions of all constructs and the relations between them. Such a network aims to provide a verifiable theoretical framework. The lack of such a network is one of the most pregnant problems affective computing is coping with. Kia Höök describes emotions as if we can pinpoint them. Although intuitively this is indeed the case, in practice it proves to be very hard to define emotions (Duffy, 1941; Kleinginna & Kleinginna, 1981).

Par excellence, humans can recognize patterns in noisy environments. Moreover, the ease with which humans adapt to new situations, to new patterns remains striking. Moreover, this is in sharp contrast with the performance of signal processing and pattern recognition algorithms. Often, these perform well in a controlled environment; however, in the “real world” their performance deteriorates (Healey, 2008). This problem refers to the influence of the context on measurements, which is also denoted as ecological validity. Due to a lack of real world research, in general, the ecological validity of research on affective computing is limited and its use often still has to be shown in “real world” practice. However, as Kia Höök illustrates, some nice exceptions to this statement have been presented throughout the last decade.

In 1941, Elizabeth Duffy published her article “An explanation of ‘emotional’ phenomena without the use of the concept ‘emotion’” in which she starts by stating that she considers *"... 'emotion', as a scientific concept, is worse than useless. ... 'Emotion' apparently did not represent a separate and distinguishable condition."* Although this statement is 60 years old it is still (or, again) up to date, perhaps even more than ever (cf. Kleinginna & Kleinginna, 1981). Almost fifty years later, in 1990, John T. Cacioppo and Louis G. Tassinary expressed a similar concern; however, they more generally addressed the complexity of psychophysiological relations. These *"are conceptualized in terms of their specificity (e.g., one-to-one versus many-to-one) and their generality (e.g., situation or person specific versus cross-situational and pancultural)."* (Cacioppo & Tassinary, 1990). They proposed a model, which *"yields four classes of psychophysiological relations: (a) outcomes, (b) concomitants, (c) markers, and (d) invariants."* Although Cacioppo and Tassinary (1990) discuss the influence of context, they do not operationalize it; hence, this discussion's value for affective computing is limited. Nevertheless, articles such as this are food for thought. Regrettably, attempts such as this are rare in the community of affective computing; consequently, the field's research methods are fragile and a solid theoretical framework is missing (Van den Broek, 2011).

To ensure sufficient advancement, it has been proposed to develop computing entities that respond on their user(s) physiological response(s), without the use of any interpretation of them in terms of emotions or cognitive processes (Tractinsky, 2004). This approach has been shown to be feasible for several areas of application. However, this approach also undermines the position of affective computing itself as a field of research. It suggests that emotion research has to mature further before affective computing can be brought to practice. This would be an honest conclusion but a crude one for the field of affective computing. It implies that affective computing should take a few steps back before making its leap forward. A good starting point for this process would be the hot topics on emotion research that Gross (2010, p. 215) summarized in his article "The future's so bright, I gotta wear shades" (see also Van den Broek, 2011).

Taken together, Kia Höök should be acknowledged for her concise overview of emotion in technology. In her chapter she takes the affective interaction standpoint. In contrast, with this commentary, I have taken an affective computing standpoint. Moreover, Kia Höök has taken a design perspective, where this commentary touches upon the fundaments of emotions in technology. I pose that if anything, affective computing has to learn more about its roots (e.g., Arnold, 1968; Candland, 1962; Dunbar, 1954); then, affective computing can and probably will have a bright future!

## References


- Arnold, M.B. (1968). *The nature of emotion: Selected readings*. Harmondsworth, Middlesex, England: Penguin Books Ltd.
- Brosschot, J.F. (2010). Markers of chronic stress: Prolonged physiological activation and (un)conscious perseverative cognition. *Neuroscience & Biobehavioral Reviews, 35(1)*, 46-50.
- Cacioppo, J.T. and L. Tassinary, L.G. (1990). Inferring psychological significance from physiological signals. *American Psychologist, 45(1)*, 16-28.
- Candland, D.K. (1962). *Emotion: Bodily change - An enduring problem in psychology, Selected readings*. Princeton, NJ, USA: D. van Nostrand Company, Inc.
- Duffy, E. (1941). An explanation of “emotional” phenomena without the use of the concept “emotion”. *Journal of General Psychology, 25*, 283-293.
- Dunbar, F. (1954). *Emotions and bodily changes: A survey of literature on psychosomatic interrelationships 1910—1953* (4th ed.). New York, NY, USA: Columbia University Press.

- Fisher, R.A. (1918). The correlation between relatives on the supposition of Mendelian inheritance. *Transactions of the Royal Society of Edinburgh, 52(2)*, 399-433.
- Gross, J.J. (2010). The future's so bright, I gotta wear shades. *Emotion Review, 2(3)*, 212-216.
- Healey, J.A. (2008). Sensing affective experience. In J.H.D.M. Westerink, M. Ouwerkerk, T. Overbeek, W.F. Pasveer, and B. de Ruyter (Eds.), *Probing Experience: From Academic Research to Commercial Propositions* (Part II: Probing in order to feed back), Chapter 8, p. 91-100. Series: Philips Research Book Series, Vol. 8. Dordrecht, The Netherlands: Springer Science + Business Media B.V.
- Kleinginna, P.R. and Kleinginna, A.M. (1981). A categorized list of emotion definitions, with a suggestion for a consensual definition. *Motivation and Emotion, 5(4)*, 345-379.
- Picard, R.W. (1995). *Affective Computing*. Technical Report No. 321. Perceptual Computing Section, M.I.T. Media Laboratory, Cambridge, MA, USA.
- Tractinsky, N. (2004). Tools over solutions? Comments on Interacting with Computers special issue on affective computing. *Interacting with Computers, 16(4)*, 751-757.
- Van den Broek, E.L. (2011). *Affective Signal Processing (ASP): Unraveling the mystery of emotions*. PhD-thesis, Human Media Interaction (HMI), Faculty of Electrical Engineering, Mathematics, and Computer Science, University of Twente, Enschede, The Netherlands.
- Watanabe, S. (1969). *Methodologies of Pattern Recognition*. New York, NY, USA: Academic Press, Inc.

## 12.8 COMMENTARY BY JOYCE H. D. M. WESTERINK

How to cite this commentary in your report

### Joyce H. D. M. Westerink

*© Joyce H. D. M. Westerink*

Joyce H.D.M. Westerink (1960) studied physics and took her Ph.D. in 1991 on the human-oriented topic of perceived image quality. She joined Philips Research and specialized on human perception, emotion and cognition related to consumer products. Written output of her work can be found in some 50 articles in books and international journals and 20 patents and patent applications....

Joyce H. D. M. Westerink
Joyce H. D. M. Westerink is a member of The Interaction Design Foundation

Kristina Höök has given us an inspiring view of three directions of research targeted at the crossroads of technology and affect, namely (traditional) Affective Computing, Affective Interaction, and Technology as Experience. She emphasizes that each line of research has contributed to the development of applications for various types of users, since they are complementary in their approach. I can only underline this conclusion from my experiences in industrial research. A few aspects in particular I'd like to single out for further discussion.

Let me start with an assumption that is contained in this and many other texts and views on Affective Computing, but never stated explicitly, namely that for any viable application in this domain, *you need a measurement of an emotion-relevant signal*. This could be a camera signal, as in Affector, movement signals as in eMoto, or physiological signals as in the Affective Diary. Also much of our own effort has been spent in the pursuit of unobtrusive measurement techniques for emotion-related signals, like our skin conductance wristband (Ouwerkerk, 2011). However, to reach the goal of 'making a machine that deliberately ... influences emotion or other affective phenomena', measurement is not strictly needed. A case in point is any TV-set or MP3 player: we use them all the time to change our mood with music, or have a TV-show experience that propels us through a series of emotions. That it works is because people are similar in their reactions to a certain extent and because TV-show directors and music composers are very skilled in creating emotional experiences for the general audience, or for specific target groups. Nevertheless, in our domain of research, everyone tacitly assumes that measurement of emotion-related signals is necessary, and indeed it allows for a further refinement of the affective influencing. Especially for changes away from the average of the crowd, in the direction of adaptation to individuals. This means that ultimately, *individual models* are not only necessary in the Affective Interactional approach, as Kia Höök proposes, but also in the Affective Computing paradigm. With the emotion-related measurements aboard, we also immediately enter the domain of *closed-loop applications* (see Van Gerven et al., 2009, Van den Broek, 2011): the emotion-related measurements are interpreted in terms of affect, then a decision is made what actions are applicable (based on present and previous measurements), and these actions are executed, after which a new measurement is done to check the new situation, etc. etc. (see Figure 1). The closed-loop model basically describes that whenever there are measurement data available, they are used to try and achieve a better situation. In this way, one's (affective) state can be guided in a targeted direction. Our Affective Music Player (Janssen et al., 2011, Van der Zwaag et al., 2009), constructed in the best Affective-Computing tradition, can serve as an example: it measures my personal reactions to music, and uses this information to adapt the playlist to direct me (not others) to a certain chosen target mood. All in all, I conclude that emotion-related measurements, individual models, and closed-loop applications are tightly interlinked in any research line in our domain.

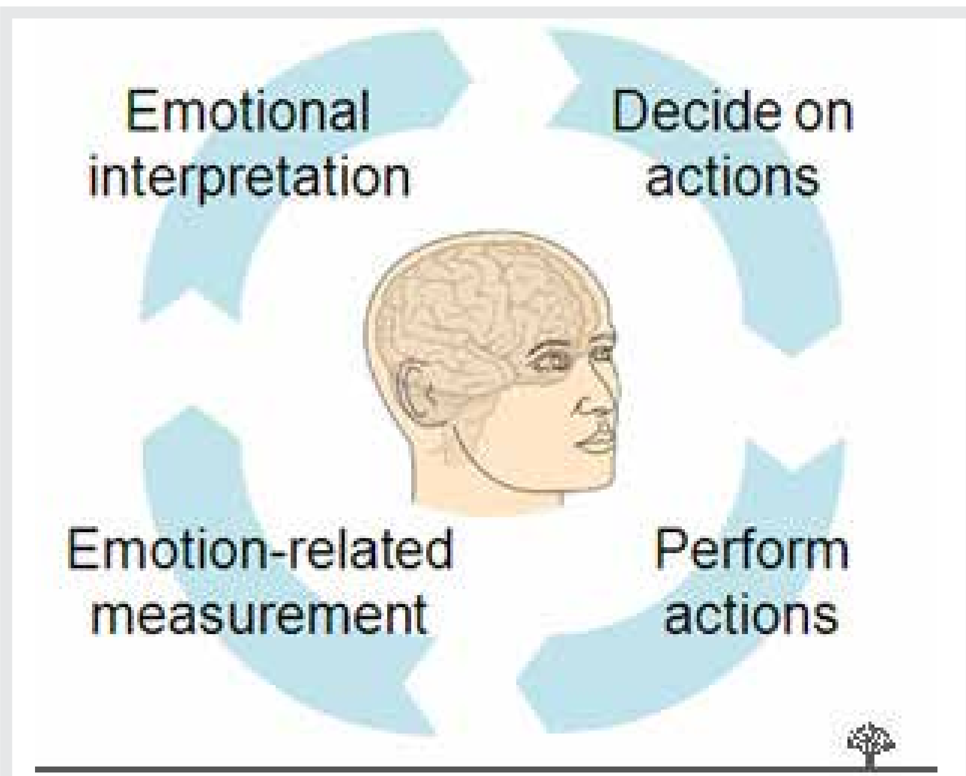



Figure 12.1: Emotional Closed Loop.

The affective closed loop in Figure 1 reserves a substantial part for *interpretation of the emotion-related signal*. This interpretation can be done *by a human*, as Kia Höök advocates along the lines of the Affective Interaction paradigm, and this human can either be the person that is measured (e.g. Affective Diary) or someone else (e.g. Affector). In both cases, the measurement information will be used to reflect on the situation measured, and if needed, to take action to change it (making it a closed loop indeed). If the raw emotion-related signals are presented, we will not stand the chance to lose information that is of value to the user, that is true. But on the other hand, this information might also be overwhelming (at least at first) and a user could benefit from help in the form of an interpretation made *by an algorithm* (in the Affecting Computing tradition). There is no need to make a choice between the two alternatives, we could think of implementing both. For instance, our electronic wristband does show the raw skin conductance/arousal patterns over the course of a day or week, but we can also give the user a discreet buzz (vibration alarm) whenever an algorithm interprets that tension has risen considerably. Of course, Kia Höök points out that it is difficult to make the correct interpretation as *context is varying* in many applications, and this is underlined by the fact that much of the research effort in affective computing has gone into algorithms deriving affective states from emotion-related signals. Nevertheless, there are options to try and overcome this: a technological approach is by adding additional sensors to monitor the context, like the accelerometer in our wristband that helps us estimate the activity level of the wearer and with that interpret the skin conductance signal. And another way out is by averaging over multiple measurements in varying circumstances to distil an overall effect. This is for instance done in our Affective Music Player, where the mood impact of a single song is modeled by taking the average affective effect (corrected for the Law of Initial Values) of multiple presentations, and this is proven to be good enough to select songs capable of directing one's mood to a certain state. Moreover, neither the raw emotion-related signal, nor its interpretation is presented to the user of our Affective Music Player: (s)he doesn't want to bother and only experiences that (s)he is brought into a different mood. Concluding, we find that both human and algorithm interpretation of emotion-related signals are important ingredients of future applications, and both are capable to deal with context to some extent.

Kia Höök argues that in normal life, emotions are always *part of a larger experience*, and that it is this larger experience that we need to support with our affective technology, in line with the Technology-as-Experience direction of research. This will certainly broaden the field of applications to include related fields in which emotions play a role. For example, emotions are important in communication, and building up relationships, and it is foreseeable that affective technologies can help (Janssen et al., 2010). It relates to the 'decide on actions' part of the closed loop in Figure 1: what do we want to do with the information gained? Nevertheless, the broadness of possible goals does not preclude that there are also applications that do have *the goal to impact affect itself*. A case in point is the Affective Music Player described before, which is exactly intended to direct affect, namely the mood. We have also shown (Van der Zwaag et al., 2011) that the optimal, individually selected, music can indeed help to *prevent the emotion (or affective state) of anger in the frustrating traffic situations* Kia Höök describes. On the other hand, I am not so sure whether consumers are interested in knowing or influencing their emotions. Despite the abundance of emotion-overloaded reality shows on TV, and despite the fact that emotion as a research topic has become fashionable in recent years, the general public still maintains a 'nice for others, not for me' attitude. In my view, this is related to the *emotion/female versus rationality/male distinction* Kia Höök mentions: The average male continues to see emotions as a female sign of weakness, of which they do not want to be reminded, not even if our measurement technology gives it a more masculine twist. For the females, it is the other way round: They do feel (more) comfortable with mood and emotions and acknowledge their impact on our everyday life, but they are less inclined to deploy masculine technology to alter them. For this reason also, I agree with Kia Höök , that our affective technologies are most likely to be used in applications that target a broader experience than that of affect alone.

To wrap up, let me highlight what I think is the most important message in Kia Höök 's story: That affective technologies will benefit from individual models (not only for human, but also for algorithm interpretation of the emotion-relate signals measured), and that they can be deployed in a wide range of applications extending far beyond the original domain of measuring and influencing affect. I am looking forward to see them appear incorporated in products and applications in the world around us....

## References

- Janssen, Bailenson, IJsselsteijn, & Westerink (2010), Intimate heartbeats: Opportunities for affective communication technology. IEEE Transactions on Affective Computing, Vol. 1, No 2, 72-80.
- Janssen, Van den Broek, Westerink, Tune in to your emotions: A robust personalized affective music player, User Modeling and User-Adaptive Interaction, In Press.
- Ouwerkerk (2011), Unobtrusive Emotions Sensing in Daily Life, in: Sensing Emotions. The Impact of Context on Experience Measurements (2011), Westerink, Krans, Ouwerkerk (eds.). Philips Research Book Series, volume 12, Springer, Dordrecht, The Netherlands, pages 21-39.
- Van den Broek (2011), Affective Signal Processing, Unraveling the mystery of emotions, Ph.D. Thesis University of Twente.
- Van der Zwaag, Westerink, Van den Broek (2009), Deploying music characteristics for an affective music player, Proceedings VOLUME I, 2009 International Conference on Affective Comput-

ing & Intelligent Interaction, ACII 2009, September 10-12, 2009, Amsterdam, The Netherlands, pages 459-465.

- Van der Zwaag, Fairclough, Spiridon, Westerink (2011). The impact of music on affect during anger inducing drives. In S. Dmello et al. (Eds.), 4th international conference on affective computing and intelligent interaction. Part I, LNCS 6974 (pp. 407-416). Memphis, Tennessee, USA: Springer, Heidelberg.
- Van Gerven, Farquhar, Schaefer, Vlek, Geuze, Nijholt, Ramsey, Haselager, Vuurpijl, Gielen and Desain (2009). The brain–computer interface cycle. Journal of Neural Engineering, Vol.6, No.4, 1-10.

## 12.4 BEHIND THE SCENES

Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.

Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.

## 12.9 REFERENCES

*Boal, Augusto (1992): Games for Actors and Non-Actors. Routledge*

*Boehner, Kirsten, dePaula, Rogerio, Dourish, Paul and Sengers, Phoebe (2007): How emotion is made and measured. In International Journal of Human-Computer Studies, 65 (4) pp. 275-291*

*Boehner, Kirsten, dePaula, Rogerio, Dourish, Paul and Sengers, Phoebe (2005): Affect: from information to interaction. In: Bertelsen, Olav W., Bouvin, Niels Olof, Krogh, Peter Gall and*

*Kyng, Morten (eds.) Proceedings of the 4th Decennial Conference on Critical Computing 2005 August 20-24, 2005, Aarhus, Denmark. pp. 59-68*

*Cañamero, Lola (2005): Emotion understanding from the perspective of autonomous robots research. In Neural Networks, 18 (4) pp. 445-455*

*Damasio, Antonio R. (1995): Descartes' Error: Emotion, Reason, and the Human Brain. Harper Perennial*

*Darwin, Charles (0000): The Expression of the Emotions in Man and Animals. London, UK, John Murray*

*Davidson, Richard J., Scherer, Klaus R. and Goldsmith, H. Hill (2002b): Handbook of Affective Sciences. Oxford University Press, USA*

*Davidson, Richard J., Pizzagalli, Diego, Nitschke, Jack B. and Kalin, Ned H. (2002a): Parsing the subcomponents of emotion and disorders of emotion: perspectives from affective neuroscience. In: Davidson, Richard J., Scherer, Klaus R. and Goldsmith, H. Hill (eds.). "Handbook of Affective Sciences". Oxford University Press, USA*

*dePaula, Rogerio and Dourish, Paul (2005): Cognitive and Cultural Views of Emotions. In: Proceedings of the Human Computer Interaction Consortium Winter Meeting 2005, Douglas, CO, USA.*

*Dewey, John (1934): Art as Experience. Perigee Trade*

*Dunbar, Robin (1997): Grooming, Gossip, and the Evolution of Language. Harvard University Press*

*Dunbar, Robin (1998): Grooming, Gossip, and the Evolution of Language. Harvard University Press*

*Ellsworth, Phoebe C. and Scherer, Klaus R. (2003): Appraisal processes in emotion. In: Davidson, Richard J., Sherer, Klaus R. and Goldsmith, H. Hill (eds.). "Handbook of Affective Sciences". Oxford University Press, USA*

*Ferreira, Pedro and Höök, Kristina (2011): Bodily Orientations around Mobiles: Lessons learnt in Vanuatu. In: Proceedings of the ACM CHI Conference on Human Factors in Computing Systems 7-12 May, 2011, Vancouver, Canada.*

*Ferreira, Pedro, Sanches, Pedro, Höök, Kristina and Jaensson, Tove (2008): License to chill!: how to empower users to cope with stress. In: Proceedings of the Fifth Nordic Conference on Human-Computer Interaction 2008. pp. 123-132*

*Gaver, William (2009): Designing for emotion (among other things). In Philosophical Transactions of the Royal Society, 364 (1535) pp. 3597-3604*

Grosz, Elizabeth (1994): Volatile Bodies: Toward a Corporeal Feminism (Theories of Representation and Difference). Indiana University Press

Hummels, Caroline, Overbeeke, Kees and Klooster, Sietske (2007): Move to get moved: a search for methods, tools and knowledge to design for expressive and rich movement-based interaction. In Personal and Ubiquitous Computing, 11 (8) pp. 677-690

Hutchins, Edwin (1995): Cognition in the wild. Cambridge, Mass, MIT Press

Höök, Kristina (2009): Affective loop experiences: designing for interactional embodiment. In Philosophical Transactions of the Royal Society, 364 p. 3585–3595

Höök, Kristina (2008): Affective Loop Experiences - What Are They?. In: Oinas-Kukkonen, Harri, Hasle, Per F. V., Harjumaa, Marja, Segerståhl, Katarina and Øhrstrøm, Peter (eds.) PERSUASIVE 2008 - Persuasive Technology, Third International Conference June 4-6, 2008, Oulu, Finland. pp. 1-12

Höök, Kristina (2006): Designing familiar open surfaces. In: Proceedings of the Fourth Nordic Conference on Human-Computer Interaction 2006. pp. 242-251

Höök, Kristina (2010): Transferring qualities from horseback riding to design. In: Proceedings of the Sixth Nordic Conference on Human-Computer Interaction 2010. pp. 226-235

Höök, Kristina, Ståhl, Anna, Sundström, Petra and Laaksolaahti, Jarmo (2008): Interactional empowerment. In: Proceedings of ACM CHI 2008 Conference on Human Factors in Computing Systems April 5-10, 2008. pp. 647-656

Isbister, Katherine and Höök, Kristina (2009): On being supple: in search of rigor without rigidity in meeting new design and evaluation challenges for HCI practitioners. In: Proceedings of ACM CHI 2009 Conference on Human Factors in Computing Systems 2009. pp. 2233-2242

Katz, Jack (2001): How Emotions Work. University of Chicago Press

Katz, Jack (1999): How Emotions Work. University of Chicago Press

Kaye, Joseph Jofish (2006): I just clicked to say I love you: rich evaluations of minimal communication. In: Olson, Gary M. and Jeffries, Robin (eds.) Extended Abstracts Proceedings of the 2006 Conference on Human Factors in Computing Systems April 22-27, 2006, Montréal, Québec, Canada. pp. 363-368

Kort, Barry, Reilly, Rob and Picard, Rosalind W. (2001): An Affective Model of Interplay between Emotions and Learning: Reengineering Educational Pedagogy - Building a Learning Companion. In: ICALT 2001 2001. pp. 43-48

*Laban, Rudolf von and Lawrence, F. C. (1974): Effort: economy in body movement. Plays, inc*

*Ledoux, Joseph (1996): The Emotional Brain: The mysterious underpinnings of emotional life. Simon and Schuster*

*Ledoux, Joseph (1998): The Emotional Brain: The mysterious underpinnings of emotional life. Simon and Schuster*

*Longo, Giuseppe O. (2003): Body and Technology: Continuity or Discontinuity?. In: Fortunati, Leopoldina, Katz, James E. and Riccini, Raimonda (eds.). "Mediating the Human Body: Technology, Communication, and Fashion". Routledge*

*Lutz, Catherine (1986): Emotion, Thought, and Estrangement: Emotion as a Cultural Category. In Cultural Anthropology, 1 (3) pp. 287-309*

*Lutz, Catherine A. (1988): Unnatural Emotions: Everyday Sentiments on a Micronesian Atoll and Their Challenge to Western Theory. University of Chicago Press*

*McCarthy, John and Wright, Peter (2004): Technology as Experience. The MIT Press*

*Merleau-Ponty, Maurice (1958): Phenomenology of Perception. London, England, Routledge*

*Moen, Jin (2006). KinAesthetic Movement Interaction : Designing for the Pleasure of Motion (Doctoral Thesis). KTH*

*Norman, Donald A. (2004): Emotional Design: Why We Love (Or Hate) Everyday Things. Basic Books*

*Ortony, Andrew, Clore, Gerald L. and Collins, Allan (1988): The Cognitive Structure of Emotions. Cambridge University Press*

*Ortony, Andrew, Clore, Gerald L. and Collins, Allan (1990): The Cognitive Structure of Emotions. Cambridge University Press*

*Parkinson, B. (1996): Emotions are social. In British Journal of Psychology, 87 p. 663–683*

*Picard, Rosalind W. (1997): Affective computing. Ma, USA, The MIT Press*

*Purpura, Stephen, Schwanda, Victoria, Williams, Kaiton, Stubler, William and Sengers, Phoebe (2011): Fit4life: the design of a persuasive technology promoting healthy behavior and ideal weight. In: Proceedings of ACM CHI 2011 Conference on Human Factors in Computing Systems 2011. pp. 423-432*

*Russell, James A. (1980): Circumplex Model of Affect. In Journal of Personality and Social Psychology, 39 (6) pp. 1161-1178*

*Sanches*, Pedro, *Höök*, Kristina, *Vaara*, Elsa, *Weymann*, Claus, *Bylund*, Markus, *Ferreira*, Pedro, *Peira*, Nathalie and *Sjölinder*, *Marie (2010): Mind the body!: designing a mobile stress management application encouraging personal reflection. In: Proceedings of DIS10 Designing Interactive Systems 2010. pp. 47-56*

*Schiphorst, Thecla (2007): Really, really small: the palpability of the invisible. In: Proceedings of the 2007 Conference on Creativity and Cognition 2007, Washington DC, USA. pp. 7-16*

*Sengers*, Phoebe, *Boehner*, Kirsten, *Warner*, Simeon and *Jenkins*, *Tom (2005): Evaluating Affector: Co-Interpreting What 'Works'. In: CHI 2005 Workshop on Innovative Approaches to Evaluating Affective Systems 2005.*

*Sheets-Johnstone, Maxine (2009): The Corporeal Turn: An Interdisciplinary Reader. Imprint Academic*

*Sheets-Johnstone, M. (1999): Emotion and Movement: A beginning Empirical-Phenomenological Analysis of Their Relationship. In Journal of Consciousness Studies, 6 (11) pp. 259-277*

*Shusterman, Richard (2008): Body Consciousness: A Philosophy of Mindfulness and Somaesthetics. Cambridge University Press*

*Ståhl*, Anna, *Höök*, Kristina, *Svensson*, Martin, *Taylor*, Alex S. and *Combetto*, Marco *(2009): Experiencing the Affective Diary. In Personal and Ubiquitous Computing, 13 (5) pp. 365-378*

*Sundström*, Petra, *Ståhl*, Anna and *Höök*, *Kristina (2007): In situ informants exploring an emotional mobile messaging system in their everyday practice. In International Journal of Human-Computer Studies, 65 (4) pp. 388-403*

*Sundström*, Petra, *Jaensson*, Tove, *Höök*, Kristina and *Pommeranz*, *Alina (2009): Probing the potential of non-verbal group communication. In: GROUP09 - International Conference on Supporting Group Work 2009. pp. 351-360*

## About the author

### Kristina Höök



Has also published under the name of:
"Kristina Hook"

Personal Homepage:
http://www.sics.se/~kia

Current place of employment:
KTH

Kristina Höök is a professor in Interaction Design at Royal Institute of Technology (KTH) in Sweden. She started and now works in the Mobile Life centre. She also upholds a part-time position at SICS (Swedish Institute of Computer Science).

## Your notes and thoughts on chapter 12

Record your notes and thoughts on this chapter. If you want to share these thoughts with others online, go to the bottom of the page at:
http://www.interaction-design.org/encyclopedia/affective_computing.html

**NOTES:**

# Chapter 13

# Requirements Engineering from an HCI Perspective

***by Alistair G. Sutcliffe.***

Requirements Engineering is, as its name suggests, the engineering discipline of establishing user requirements and specifying software systems. There are many definitions of Requirements Engineering (Zave, 1995); however, they all share the idea that *requirements* involves finding out what people want from a computer system, and understanding what their needs mean in terms of design. Requirements Engineering is closely related to software engineering, which focuses more on the process of designing the system that users want. Perhaps the most concise summary comes from Barry Boehm: requirements are "designing the right thing" as opposed to software engineering's "designing the thing right" (Boehm, 1981). Nuseibeh and Easterbrook (2000) give a more comprehensive definition as "*software systems requirements engineering* (RE) is the process of discovering that purpose, by identifying stakeholders and their needs, and documenting these in a form that is amenable to analysis, communication, and subsequent implementation".

Requirements Engineering shares many concepts, techniques and concerns with Human Computer Interaction (HCI) especially user-centred design, participatory design and interaction design. However, it differs from HCI in its view of the scope of design; for example, socio-technical design is rarely mentioned in Requirements Engineering (although see Sutcliffe et al.(2005) for an exception), where the organisational and people part of a system is an explicit target of requirements and design. The other difference lies in the HCI focus on design *per se* and interaction design where user requirements are seen as part of the process of a design exploration, prototyping and evaluation dialogue with the user, rather than as the more linear requirements "specify-design-implement" process favoured in the Requirements Engineering community. However, Requirements Engineering certainly espouses iterative design, prototyping and evaluation (usually validation in requirements terminology). Furthermore, there has been a growing realisation that requirements cannot be specified without doing some design, which had led to the emergence of "architecture requirements". In this chapter I will explore the commonalities and differences between Human Computer Interaction and Requirements Engineering to reflect on how the two communities tackle what is essentially the same problem: building a system based on software that satisfies people's needs.

The chapter is structured in six subsequent sections. In the next section, 13.1, the Requirements Engineering process is described. This is followed in section 13.2 by a review of scenario-based approaches which illustrate the convergence between Requirements Engineering and HCI. Section 13.3 deals with models and representations in the two disciplines, then section 13.4 returns to a process theme to assess the differences between HCI and Requirements Engineering approaches to development. Section 13.5 reviews how knowledge is reused in the requirements and design process, leading to a brief discussion of the prospects for convergence between HCI and Requirements Engineering.

# 13.1 REQUIREMENTS ENGINEERING ACTIVITIES AND PROCESS

## 13.1.1 Scoping

Requirements frequently start with a vague statement of intent. The first problem is to establish the boundary of investigation and, *inter alia*, the scope of the intended system. Unfortunately, this is rarely an easy process as clients often don't know exactly what they want, and knowledge about the intended system is vague. Scoping tends to be an iterative activity as the boundaries become clearer with increasing understanding of the domain shared by all the stakeholders. However, the process is poorly understood and little research has directly addressed this difficult problem.

Take for example a system to help epidemiologists with their research, which was the brief for the ADVISES project (Thew et al., 2009; Sutcliffe et al., 2011). The stakeholders might include public health analysts with interests in epidemiology as well as medical researchers. The range of possible decision-support tools could include data collection, data preparation, statistical analysis, visualisations, graphs, maps, as well a groupworking support for collaborative discussion of results. The systems owner and scope were not clear since the project was initiated as part of the UK Government supported e-science research programme, with users who were academic researchers in epidemiology, and who also collaborated with public health analysts who worked in local hospitals.

For general scoping, enterprise modelling (Kirikova & Bubenko, 1994) provides a way of describing the business context to discover requirements in the large (i.e. goals, aims, policies), but little process guidance is offered. Workshops in the KJ brainstorming method, named after its inventor Jiro Kawakita, and Rapid Applications Development (DSDM Consortium, 1995) are the current state of the art; they advocate use of lists and informal maps of the problem space, although these methods also offer little systematic guidance. More detailed scoping has been researched by Jackson and Zave (1993), who propose techniques for establishing the system boundary by examination of the intended system's obligations in responding to real-world events, although this does not help bounding investigations which start from general statements of users' intentions.

Scoping is best achieved by discussion with all the stakeholders and by documenting the high-level system goals as terms of reference. Writing down the scope tends to focus users' attention on where the boundaries of the system investigation should lie, and helps to identify at least an initial scope for the system.

### 13.1.2 Fact gathering

For the most part, the techniques for this activity have been borrowed from systems analysis, e.g. interviews, observation, questionnaires, text and document analysis (Gause & Weinberg, 1989). Techniques from knowledge acquisition, such as repertory grids and protocol analysis, have been employed, but there have been no systematic investigations into the merits of different fact-capture techniques, apart from a preliminary study by Maiden and Rugg (1994). An interesting emergent area is the use of ethnographic and associated observational methods (Goguen & Linde, 1993; Luff et al., 1993); however, these have failed so far to deliver explicit guidance for fact capture or analysis, leading software engineers to propose their own quick and dirty approaches (Hughes et al., 1995).

### 13.1.3 Analysis

Analysis and modelling generally follow top-down approaches, concentrating on goal decomposition. Analysis is often driven by 5 'W' questions:

- What is the system purpose (goals)?
- What objects are involved?
- Where is the system located?

- When should things happen?
- Why is the system necessary (goals or problems it intends to solve)?

In an example of the common approach, Potts et al.(1994, 1995) provide a means of goal-related analysis which uses scenarios to discover obstacles, or potential problems caused by external agents that the system has to deal with. From obstacles, goals for maintaining, avoiding and repairing situations can be elaborated. Other goal decomposition methods follow a taxonomic approach and attempt to analyse goals in the context of domain models (Sutcliffe & Maiden, 1993). For problem analysis, soft systems methodology (Checkland, 1981) gives a means of informal modelling and an analytic approach to discovering problem-oriented requirements. Informal diagrams and sketches, which may be referred to as domain models or rich pictures (Checkland 1981) are used to document the analysis as it progresses. A domain model for the ADVISES project is illustrated in Figure 13.1.

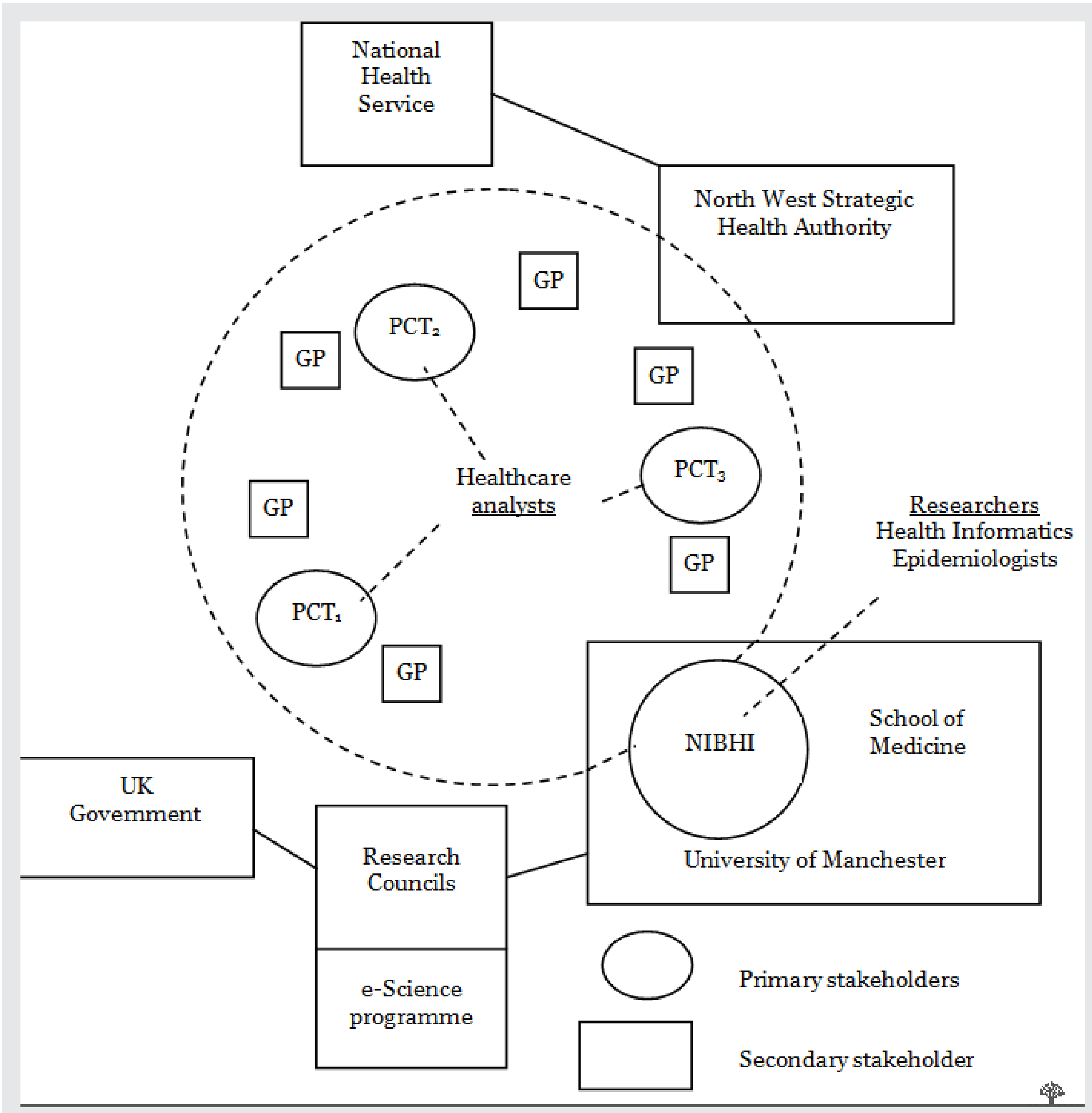



FIGURE 13.1: Domain model for the ADVISES project, as an informal diagram showing stakeholders and organisations involved.



The organisations involved appear in rectangles as secondary stakeholders, i.e. people who have an interest in the system output, but are not the primary hands-on users who are shown in circles. Primary stakeholders were public health analysts in Primary Care Trusts (local units of organisation in the UK National Health Service) and academic health informatics researchers who collectively formed the NIHBI unit (North West Institute of Bio-Health Informatics).

Rationale-based techniques are also appropriate for analysis and modelling. These structure analysis in hierarchies of graphs linking goals with potential solutions and supporting arguments; see gIBIS (Conklin & Begeman, 1988) and QOC (MacLean et al., 1991).

### 13.1.4 Modelling

This activity consumes the output from analysis, structures facts and represents them in a notation. Requirements Engineering has borrowed techniques for this activity from structured system development methods and conceptual modelling. Informal modelling notations, such as data flow diagrams and entity relationship diagrams, have been widely used. Many formal approaches to modelling have been imported from software engineering (Goguen & Linde, 1993; Van Lamsweerde, 2009; Van Lamsweerde et al., 1995) although the effectiveness of these techniques has yet to be demonstrated in industrial practice. Analysis and modelling are frequently interleaved to elaborate the requirements as understanding of the problem domain increases through the act of representation.

Some examples of modelling notations are shown in Figures 2-4. The entity relationship diagram in Figure 13.2 shows entities and their functional relationships (has, prices, etc.) which support data modelling and ultimately database design.

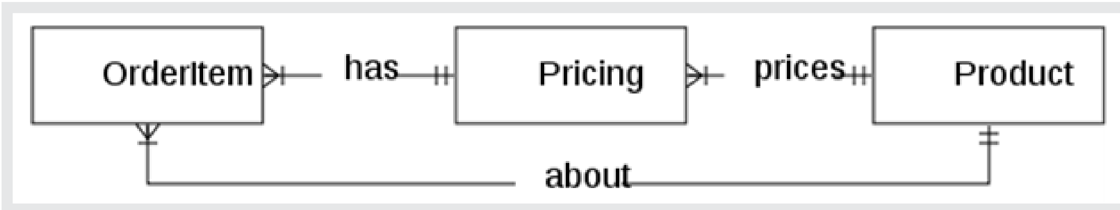



FIGURE 13.2: Entity relationship diagram; entities are shown as rectangles, relationships as connecting lines.



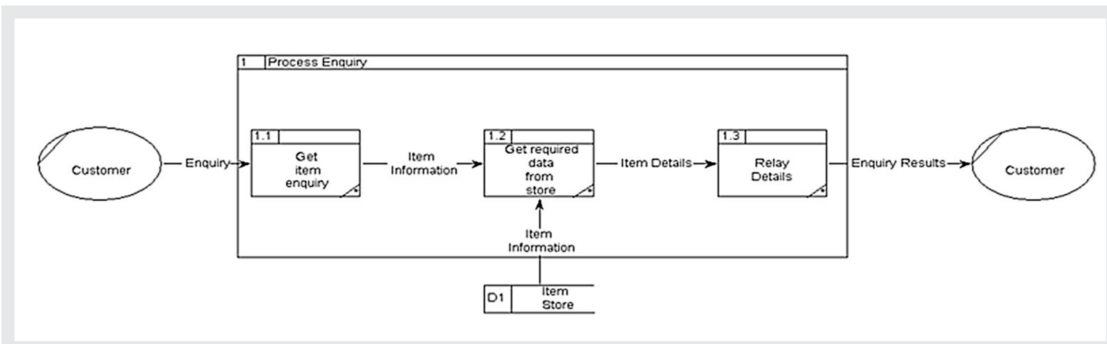



FIGURE 13.3: Data flow diagram: rectangles in this notation are processes, ovals are external agents and arrows show the directions of the flow of information between processes.



Data flow diagrams model the processing aspects of a system, and complement the "data view" model of entity relationship diagrams.

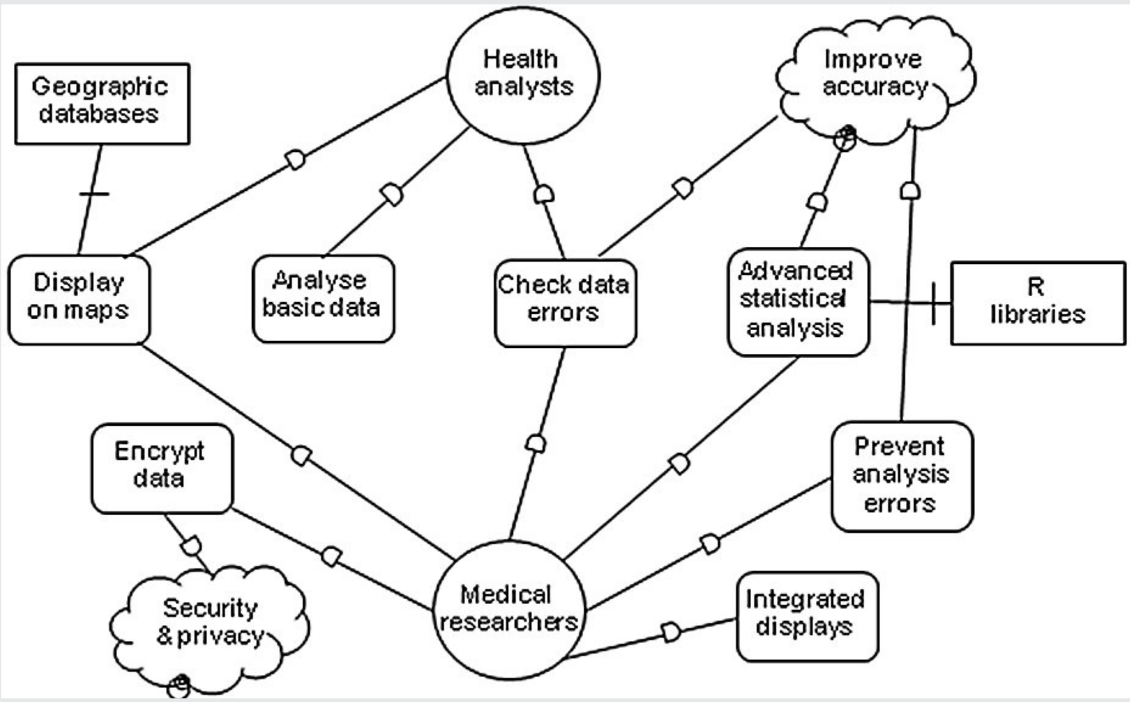



FIGURE 13.4: *i** diagram, also known as GRN (goal requirements notation).



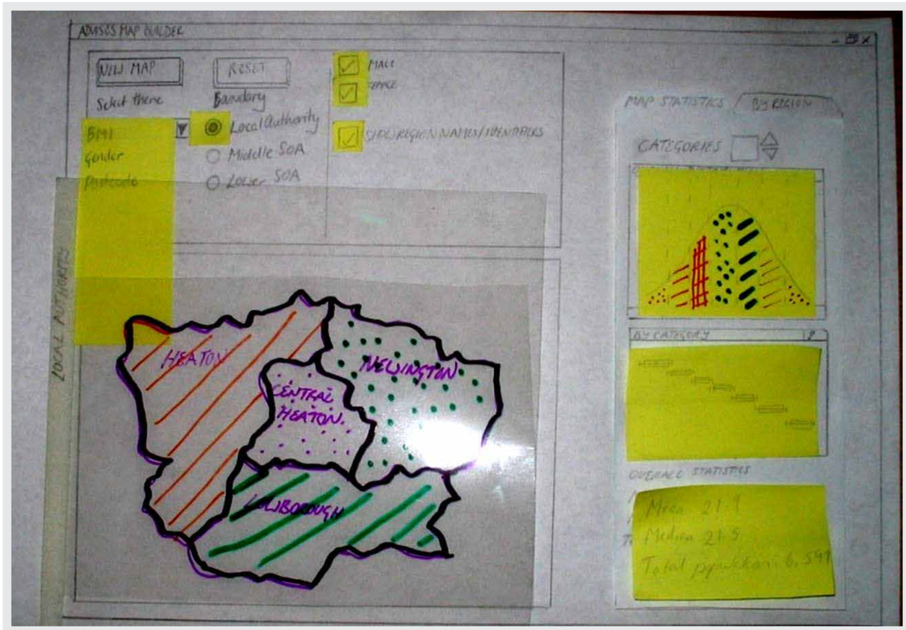
Circles represent agents, ovals are goals, cloud-like shapes denote "soft goals" or quality criteria, rectangles are resources, and the relationships with D annotation are dependencies. Tasks which achieve goals have been omitted to simplify the diagram.

The *i** diagram in Figure 13.4 illustrates the relationships between the stakeholders, major goals and non-functional requirements (soft goals in *i** terminology) in the ADVISES application. Dependencies between agents and goals are marked with Dsymbols on the arcs.

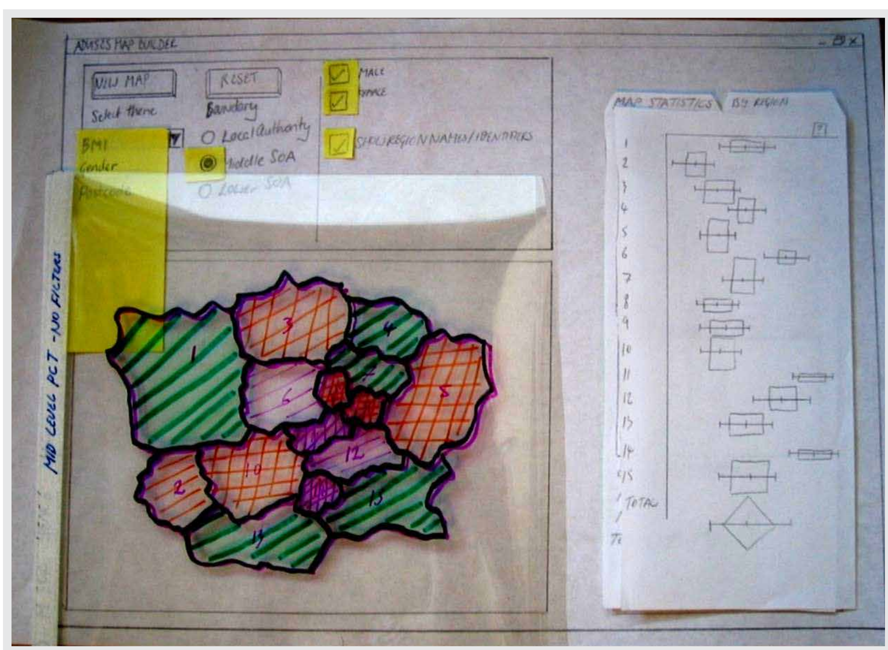
### 13.1.5 Validation

This key activity in Requirements Engineering, in spite of being extensively researched, is still problematic. Validation implies getting users to understand the implications of a requirements specification and then agree, i.e. validate, that it accurately reflects their wishes. The current state of the art is walkthrough techniques in which semi-formal specifications such as data flow diagrams are critiqued in a workshop of designers and users. Walkthroughs have the merit of early validation on specifications, whereas prototypes are probably more powerful as users react more strongly to an actual working system. Unfortunately prototypes still incur construction costs, and poorly organised use of prototyping can be detrimental (Attwood et al., 1995). However, prototypes in combination with techniques for gathering and evaluating user feedback can be highly effective (Gould, 1987). Overall, the process of validation is poorly understood and explanation is an important yet often neglected component. Some research into explaining complex requirements has demonstrated that a combination of visualisation, examples and simulation is necessary (Carroll et al., 1994; Maiden & Sutcliffe, 1994). Scenario-based representations and animated simulations help users see the implications of system behaviour and thereby improve validation (Johnson et al., 1992); furthermore, early prototypes with scenarios are a powerful means of eliciting validation feedback (Sutcliffe, 1995). The inquiry cycle technique of Potts et al.(1994) approaches validation by comparing scripts of imagined real-world behaviour against the required behaviour in a specification. Validation is still poorly understood and further research is necessary to discover how explanation, representation and users' understanding of system specifications interact.

In ADVISES, requirements validation was an iterative process of showing users different designs initially as storyboards, and later as mock-ups and prototypes. This facilitated discussion with users about just what they wanted, and also helped the Requirements Engineering team learn more about the users' work. Presenting people with designs rarely fails to elicit useful feedback, and this approach is central to user-centred design.




Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.




Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.

FIGURE 13.5 A-B: Illustrations of the ADVISES storyboards used in validation sessions with users.

A video demonstration of the ADVISES prototype can be found in the supplementary material or at: http://www.youtube.com/watch?v=8EfSM9KG3Dg

Figure 13.5 shows two configurations of the ADVISES paper prototype designed to support exploration of epidemiological data. (A) shows a map of a fictitious city split up according to Primary Care Trust boundaries, including an apparent hotspot in one area shown by shading to indicate that the mean for this area falls at the extremes of the distribution. (B) shows the same map but now split into smaller sub-areas accompanied by a histogram-like plot for those areas. Acetate transparencies are used as overlays so different options can be presented in an interactive sequence, while post-it notes are used to record ideas and feedback suggestions during validation sessions. Users are encouraged to draw on the paper prototypes to illustrate their own design ideas. Figure 13.6 illustrates an interactive requirements validation session where post-it notes profile a means of sharing ideas among all the participants.

Figure 13.6: Users discussing ideas on post-it notes during a requirements validation session.

### 13.1.6 Trade-off Analysis

Requirements frequently cannot be satisfied by a specification. Non-functional requirements fall into this category; although the design can accommodate them to some degree, a complete solution is not feasible. Requirements are often held by different stakeholders who may have conflicting views, hence trade-off analysis is an essential activity for comparing, prioritising and deciding between different requirements or design options. Ranked lists or matrix-based techniques using decision tables are helpful for this analysis. The modelling techniques proposed by Chung (1993) and Yu (1993) for mapping relationships and dependencies between goals, tasks, actors and soft goals (alias non-functional requirements), contains some guidance for trade-off analysis. Their method and support tool facilitate tracing influences between goals and non-functional requirements (NFRs) as well as giving active guidance about potential clashes between different types of NFR (e.g. security may militate against ease of use). This work is a significant advance in handling trade-offs. In spite of this, few tools or methods exist to help the requirements engineer, although House of Quality (Hauser and Clausing, 1988) techniques have been imported into RE and some tool support is available (Jacobs and Kethers, 1994). More complex approaches such as multi-criteria decision making (Fenton, 1995) do not seem to have been considered in RE.

An example of trade-off analysis in ADVISES is shown in Table 1. Two sets of requirements were gathered, one from the academic researchers and the other from the public health analysts. The relationship between the requirements and different quality criteria or NFRs, such as accuracy, privacy, etc., are shown in the table.

| Goals | Medical researchers | Health analysts | Non-functional requirements | | |
|---|---|---|---|---|---|
| | | | Accuracy | Security | Usability |
| Plot data on maps | ✓ | ✓✓ | | | ++ |
| Show hotspots on maps | - | ✓✓✓ | | | ++ |
| Provide simple statistics | ✓ | ✓✓ | | | + |
| Annotate maps | - | ✓✓✓ | | | + |
| Check data errors | ✓✓ | ✓ | ++ | + | |
| Provide advanced statistics | ✓✓✓ | - | ++ | | |
| Prevent analysis errors | ✓✓✓ | -- | ++ | | |
| Encrypt data | ✓✓ | - | | ++ | |
| Integrate maps and charts | ✓✓ | - | + | | + |

TABLE 13.1: The dependencies between requirements goals and quality criteria for two stakeholder views in ADVISES. Ticks show the prioritisation of the goals for each group of stakeholders, - are neutral, -- shows disapproval or a goal conflict, while + implies associations between goals and NFRs.

Decision tables, trees and flowcharts are other representations which can facilitate negotiation and trade-off analysis by making the space of options clear and helping stakeholders to see their priorities from the perspective of others.

### 13.1.7 Negotiation

The social dimension of Requirements Engineering is poorly understood. This activity subsumes many others, e.g. analysis, trade-off and modelling, but the essence lies in discussion, explanation and negotiation of conflicting requirements. The negotiation issues in ADVISES are illustrated in Table 1. Clearly the two groups of stakeholders have different priorities with only a few goals in common (data errors, maps and simple statistics). If stakeholder goals do not conflict, then all goals might be included in the design, although this can increase complexity. However, when goals conflict, as was the case in the "prevent analysis errors" goal, negotiation has to reconcile the conflicting views. Health analysts perceived this goal as a slight on their professional integrity; however, when the positive effects of this function on improving the accuracy of results was explained, they accepted this requirement. The modelling work of Chung (1993) contributes to negotiation by creating a shared artefact through which influences and design alternatives can be discussed. This is effected by creating a strategic dependency model to map out relationships between goals, tasks, actors, etc., followed by a strategic rationale model which illustrates potential system solutions for the requirements with arguments for and against them. Unfortunately, these models provide no active guidance for agreeing requirements, although Boehm et al. (1994) suggest some heuristics for structuring successful negotiation of requirements. Stakeholder analysis methods in co-operative requirements capture (Macaulay, 1993) help to structure the composition of workshops with different stakeholders and provide a framework for considering requirements from different viewpoints. Guidance for managing Requirements Engineering meetings, and handling negotiation and conflict resolution, is hard to find. Social science research on meetings describes roles, desiderata for leadership and managing consensus in groups (Viller et al., 1994); however, this research has not been applied in Requirements Engineering.

FIGURE 13.7: Different groups of users may have conflicting requirements which are negotiated in workshops facilitated by the designers.



It helps to have experts in the negotiation process facilitate discussion in a relaxed setting to reduce tensions and possible conflicts among stakeholders, as shown Figure 13.7.

The Requirements Engineering activities are placed into a generic process map in the following section to illustrate a typical route.

### 13.1.8 Process Map of Requirements Engineering

Requirements Engineering follows different routes and mixes of activities; for instance, requirements might start as problems with a current system, or as examples of products which users want. Procurement-based Requirements Engineering to select products is beyond the scope of this chapter (see Maiden & Ncube, 1998 for COTS-based Requirements Engineering), so only the most common, goal-oriented, pathway will be explained. Requirements are initiated by senior managers and company executives as policies, aims, objectives and other high-level statements of intent. This route, illustrated in Figure 13.8, necessitates considerable scoping activity as requirements start with vaguely expressed intentions and users' wish lists. The process borrows from business analysis, such as *Set policy objectives* (1) and *Analyse and model business* (2). Policy can be analysed within the business context by enterprise models. Non-Requirements Engineering activities which are pertinent to policy analysis include business modelling, value-chain analysis (Porter, 1980), competitive advantage theories and business process re-engineering (Davenport, 1993). Business analysis techniques such as business process analysis, concept maps (Eden, 1988), and critical success factors (Rockart & Short, 1991) are also applicable at this stage; however, proposing a detailed methodology is beyond the remit of Requirements Engineering. The key problem is to model the business to discover opportunities for developing computer systems to enhance competitive advantage. Although some suggestions can be found in value-chain models e.g. (Porter, 1980), and case histories of inter-organisational system design (Holland, 1995), this area is poorly understood. The methods and approaches in the business analysis community are still largely a matter of intuition, so regrettably only limited guidance can be gleaned from this source.

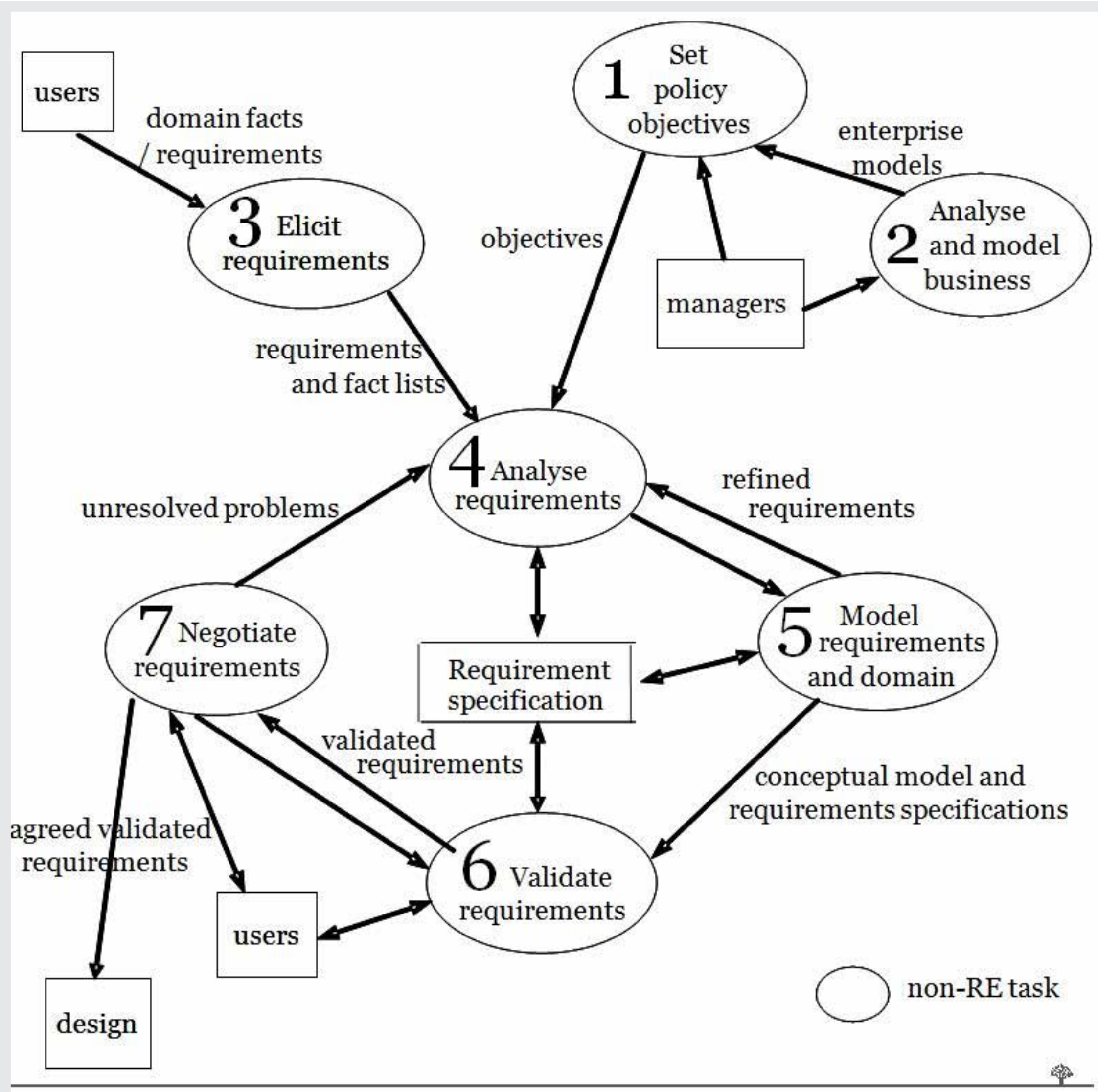



FIGURE 13.8: Goal-oriented pathway for Requirements Engineering.



Top-down decomposition is the normal approach whereby policy-level intentions are successively decomposed into goals. Objectives from management are combined with facts, information and goals from users gathered in the *Elicit requirements* (3) step via interviews, focus groups, workshops, etc. These form the inputs for *Analyse requirements* (4) in which preliminary information, usually in the form of lists and notes, is organised as connections between facts are made. For example, relationships are added progressively as the context of the policy is understood in terms of what has to be done to achieve it (goals) and the implications for people (actors) and their organisations (organisation unit, objects, etc.). *Analyse requirements* is interleaved with *Model requirements and domain* (5) as analysis produces models that document facts and their associations. For example, modelling goals in the context of how they impact on tasks and the organisation is vital not only to elaborate the meaning of informal statements of intent but also to enable assessment of the impact of change (Chung, 1993; Yu, 1993). Goals have to be refined as linguistic statements of intent until the stage when the desired state of the system can be described, when formalisation may be possible (Dardenne et al., 1993). Hypertext tools can help to represent informal goal hierarchies (Pohl, 1996), as can standard conceptual models, e.g. data flow diagrams, but there is little advice or process guidance for goal-related requirements analysis. Chung (1993) and Yu (1993) provide representations of goals in context models showing dependencies between goals (both functional and non-functional), actors and tasks, with some guidelines for goal decomposition and modelling. Modelling is an essential precursor for the *Validate requirements* stage (6) in this route, as goals can not be easily understood without contextual detail about how they may be achieved and their relationship to agents and processes. *Validate requirements* is usually interleaved with *Negotiate requirements* (7) as discussion among stakeholders and designers leads to requirements being agreed or rejected, and conflicts between stakeholders being resolved. Stages 4-7 in the process form an iterative loop as requirements are rarely specified correctly first time; instead, requirements emerge through iterations of analysis, modelling, validation and negotiation. The policy route converges with other Requirements Engineering pathways for common activities of trade-off analysis and negotiation, both of which are important for goal-oriented Requirements Engineering. Once goals are decomposed to the stage where the desired state of the system can be described, at least informally, the first-cut decisions on use of technology can be made. Some goals become functional requirements, while others have implications for management alone (e.g. decisions about resources, organisation, and human activity).

Having described Requirements Engineering activities, the next section reviews scenario-based approaches which form the common ground between the Requirements Engineering and HCI.

## 13.2 SCENARIO-BASED DESIGN: COMMON GROUND BETWEEN HCI AND RE

HCI and the parent discipline of Requirements Engineering, Software Engineering (SE), are both design disciplines that aim to develop software systems. Their close relationship has been subject to considerable debate although unfortunately little constructive synthesis. Requirements Engineering may be regarded as the front end of Software Engineering since it focuses on requirements and process phases prior to implementation, although the boundary between Requirements Engineering and Software Engineering is becoming increasingly blurred. HCI, in contrast, covers the whole design process. The boundary with Software Engineering depends on the focus of the two disciplines. Software Engineering's core concern is software, so people and systems in the socio-technical sense are minor concerns; whereas HCI focuses on people and the user interface as well as on the design of the wider socio-technical system, at the expense of software architecture. The common ground between Requirements Engineering and HCI lies in shared processes and representations, advocated in different flavours as user-centred design, scenario-based design, iterative development and agile methods.

Both HCI and Requirements Engineering use scenarios as a motivation for design, although the form and function in each field differs. Unfortunately, the term "scenario" has been abused in the literature and a large number of definitions exist (see Rolland et al., 1998). Indeed, much of the scenario literature, especially in the Software Engineering tradition (Kaindl, 1995), is in fact describing event-sequence traces through state transition models. In object-oriented design it becomes difficult to distinguish between use cases, alternative paths through use cases, and scenarios, which are just another path through a use case (Cockburn, 2001; Graham, 1996; Jacobson et al., 1992). Scenarios have several roles in design, from a “cognitive prosthesis” stimulating the designer’s imagination to narratives of system use and problems from which requirements emerge (Carroll, 2002).

In Requirements Engineering, scenarios are generally seen as inputs to modelling and are closely related to use cases and requirements elicitation, constituting the early phase of RUP (Rational Unified Process). In contrast, HCI places less emphasis on the link from scenarios to modelling; instead, scenarios and other techniques such as personas stimulate thought in the design process (Carroll, 2002). Scenarios can be related to personas which amplify narrative experiences by describing typical users. Indeed, some propose scenarios that are deliberately exceptional to provoke constructive thought (Djajadiningrat et al., 2000). Scenarios are arguably the starting point for all modelling and design; however, modelling has fallen out of favour in HCI, and task models are rarely mentioned as components of the design process. In Requirements Engineering and Software Engineering, modelling is still a mainstream activity and this illustrates a key divergence between the disciplines. See Figure 13.2.

One productive juxtaposition of scenarios and models is to use scenarios as test data to validate design models. This approach has been actively researched in the Inquiry Cycle (Potts, 1999), which recommended using scenarios as specific contexts to test the utility and acceptability of system output. By questioning the relevance of system output to a set of stakeholders and their tasks described in a scenario, the analyst can discover obstacles to achieving system requirements. Input events can be derived from scenarios to test validation routines and other functional requirements. This approach has been refined into a formal process for discovering the achievability of system goals with respect to a set of environmental states, taken from scenarios (Van Lamsweerde & Letier, 2000).

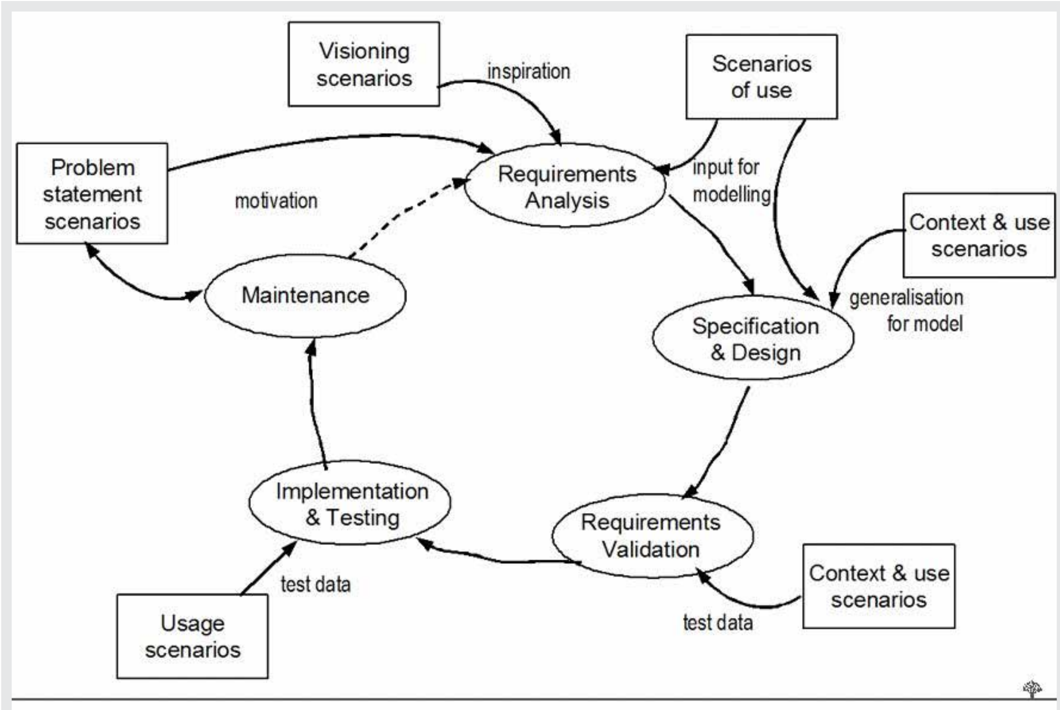



FIGURE 13.9: Use of scenarios in different phases of the Requirements Engineering-Software Engineering process.



To illustrate with some examples from the ADVISES e-science system, the initial vision scenario describes how medical researchers might work in the future:

> " Epidemiologists view data sets from different parts of the UK displayed on maps and different charts. They can ask questions about the data using limited natural language and simple controls like sliders so they can immediately see the results of different analyses on their data in different areas of the country. When they have found interesting results they can add notes and send the results to colleagues in other research teams. "

Vision scenarios set the scene for the development project before any prototype exists so they focus on the intended outcomes.

Usage scenarios, in contrast, paint a more detailed picture of how the system will operate and are frequently accompanied by storyboards and prototypes to illustrate the design:

> “ Iain wants to see if there is any link between asthma and obesity in young school children in different areas of greater Manchester. He loads data sets for the incidence of asthma and obesity in different areas. Map displays show that most areas have little association, apart from two local districts where both high levels of obesity and asthma are shown by colour coding in the map. He checks that this is an accurate result by applying an area density correction statistic and then running a correlation analysis.This shows a significant result. However, diet and poor exercise are more common causes of childhood obesity so Iain loads the location of sports facilities on the map and finds that both asthma-obesity hotspots also have few sports facilities. His investigation continues. ”

Later in the development cycle, context and use scenarios describe system use but with expectations of how the system output may be used, including test probes for use in validation and evaluation sessions, e.g.

> “ You are a public health analyst for the Greater Heaton PCT. You are interested in looking at BMI in year 6 (age 11 years) children in your area. Make a new map based on the Middle Layer Super Output Areas, and using your first data set. Load the younger children data sets and

> inspect the map display, using sliders to change the view of results using the age, gender and other variables provided. See if you can spot any general patterns when you look at maps for males only, females only and both together. ”

. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .

> “ You have noticed a hotspot in one region of your Reception Class map which appears to have higher levels of obesity. Load the younger children data sets and inspect the map display, using sliders to change the view of results using the age, gender and other variables provided. Investigate whether or not this looks like a genuine hotspot. ”

. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .

HCI uses scenarios in a similar manner in usability evaluation, although the role of scenarios is not articulated so clearly. Nevertheless, task or test scripts in evaluation methods (Monk & Wright, 1993; Sutcliffe, 2000a) are scenarios. Carroll also recognised the validation role for scenarios in the task-artefact cycle in which an implemented artefact is evaluated, leading to design improvements and, by a process of claims analysis, to new HCI knowledge. Carroll has articulated several different roles for scenarios in the design process including as envisionment for design exploration, requirements elicitation and validation (Carroll, 1995). Other roles are usage scenarios, which illustrate problems; initiating or visioning scenarios, which stimulate design of a new artefact; and projected use scenarios, which describe future use of an artefact that has been designed (Sutcliffe & Carroll, 1998).

## 13.3 MODELS AND DESIGN REPRESENTATIONS

Models are a central concern in Requirements Engineering as representations of the system and to support reasoning in the requirements process. The most popular models in Requirements Engineering are use cases, which are shared with commercial systems development and Software Engineering; see Figure 13.10.

| | |
|---|---|
| Use Case Number | EM3 |
| Name: | Viewing the name of a map region |
| Description: | Basic theme map has been created |
| Primary actor: | Epidemiologist |
| Pre-condition: | User has successfully created a theme map, divided into geographic regions |
| Trigger: | Map has been created, user is ready to explore map |
| Basic flow: | Selection of a region, viewing of name |
| Alternate flows: | Error handling – name cannot be displayed for the selected region<br>• User selects the region of the map they are interested in by hovering their mouse pointer over the region of interest<br>• After 1 second pause, hovertext display of the name of the region – this hovertext persists until the user moves their mouse |

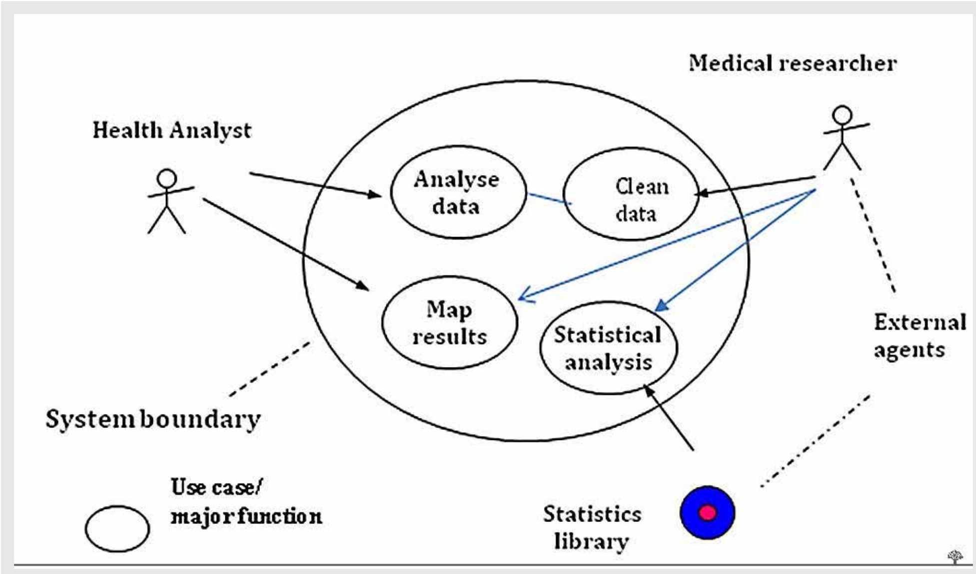



FIGURE 13.10: Use case context diagram and lower-level use case specified in an action sequence template.



More specific to Requirements Engineering is the *i** family of models which records agents, goals, tasks and resources connected by dependency and means-ends relationships, as illustrated in Figures 4 and11.

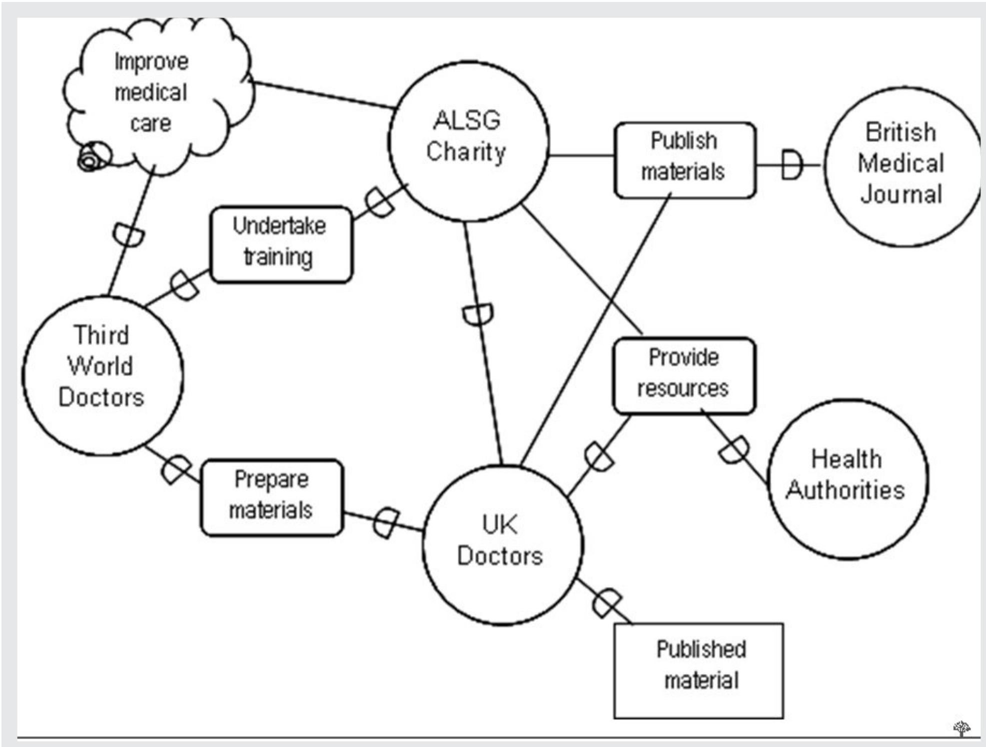



Figure 13.11: *i** strategic dependency model, showing agents (circles), goals or high-level functional requirements (rounded rectangles), resources, (rectangles) and soft goals or non-functional requirements (clouds).



Goal models and goal-oriented analysis are also key influences in Requirements Engineering (GORE: see Potts, 1999). Goal hierarchies represent the decomposition of user needs with relationships showing interactions between them such as support, inhibit or hinder. Goal models share a common heritage with task models, although task models record not only intent, but also the operational sequence in which a task will be carried out. Task models (e.g. HTA: Annett, 1996; TAKD: Diaper & Johnson, 1989) link goal-decomposition hierarchies with action sequence descriptions at lower levels of detail. Both goal and task models use a hierarchical notation, as illustrated in Figure 13.12.

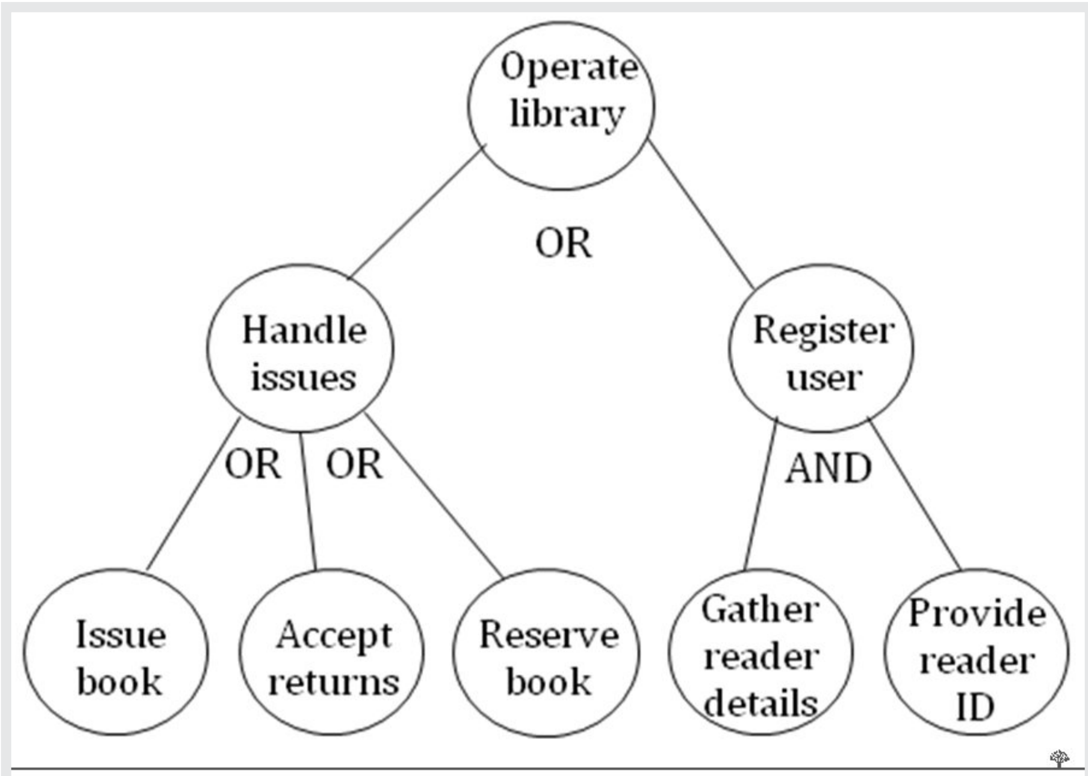



FIGURE 13.12: Library goal model with upper-level goals being decomposed into lower-level sub-goals. Task models in HCI follow a similar pattern.



Dependency relationships, denoted by the D symbol on an arc, signify an association such as a goal depends on an agent to realise it or a goal depends on a resource for its achievement. Means-end relationships record how goals will be delivered by tasks, agents and resources. Goals belong to users and are equivalent to requirements. In *i** a distinction is made between hard goals which will become functional requirements to be implemented by software programs; and soft goals (NFRs) which are quality requirements such as usability, safety, cost or accuracy.

SE extends Requirements Engineering models to provide different views of the software system, such as the data process structure in class diagrams, behaviour in activity sequence diagrams, or event sequences in state transition diagrams.

Task models provide an abstract view of the real world with a set of semantics motivated by the modelling goal. In a similar manner, class diagrams in Software Engineering represent the abstract inheritance structure of system objects, whereas state transition diagrams represent activity-oriented specification. Task models fit within the modelling genre that represents intent (goals) and activity (procedures or action sequences). Task modelling has been extended to cover data-oriented views via domain knowledge structures in TKS (Task Knowledge Structures: Johnson et al., 1988), and have been adapted to the object-oriented paradigm in MAD (Methode Analytique et Description: Rodriquez & Scapin, 1997). A prime role of task models has been to represent the problem space when reasoning about functional allocation. Although the functional allocation literature acknowledges the need for task analysis, it rarely represents task models explicitly (see Dearden et al., 2000). Instead, scenarios may be used to motivate functional allocation decisions, within task-related frameworks such as IDAS (Information, Decision, Action, Supervision: Wright et al., 2000). However, task models can be used to partition activity during functional allocation (Sutcliffe, 1997; Vicente, 2000), and this remains one of their major roles.

One criticism of task analysis is that it does not capture the richness of interaction that occurs in the real world compared with scenario narratives that concentrate on contextual description (e.g. Kuutti, 1995; Kyng, 1995). However, task models have been extended to describe information requirements implied in tasks (Sutcliffe, 1997), and the role of work artefacts in ecological interface design (Vicente, 2000). Another omission is the lack of explicit representation of communication between agents engaged in collaborative tasks, although this is partially specified in GTA (Van Der Veer et al., 1996) and is dealt with more explicitly in coupling analysis, which provides a discourse analysis framework that can be incorporated into a task workload analysis (Sutcliffe, 2000b, 2002b). Finally, task models may be criticised for not representing the relationships between agents, activity and organisational structures, although these concepts are described in socio-technical system design frameworks such as ORDIT (Eason et al., 1996); while more comprehensive modelling languages can be found in Requirements Engineering such as *i** that analyses the dependencies between agents, tasks, goals and resources (Mylopoulos et al., 1999; Yu, 1993).

Task analysis as a method or notation has not been readily adopted in practice (Bellotti, 1988; Diaper, 1999) apart from the human factors safety engineering community (e.g. Hierarchical Task Analysis: Annett, 1996). Furthermore, modelling has fallen out of favour in HCI where lightweight representations such as hierarchy diagrams for information architecture and design sketches, wireframes and storyboards have become the norm.

A shared genre of representations has been Design Rationale which has been adopted by both HCI and Requirements Engineering to record design decisions and present alternative design solutions for debate. In Requirements Engineering the gIBIS (Conklin & Begeman, 1988) notation has been adopted to represent user goals as issues, mapped to design alternatives (requirements) with arguments. QOC (MacLean & McKerlie, 1995), the HCI version, follows a similar format with Questions as design issues, Options for design alternative, and Criteria to select trade-offs between designs; see Figure 13.13.

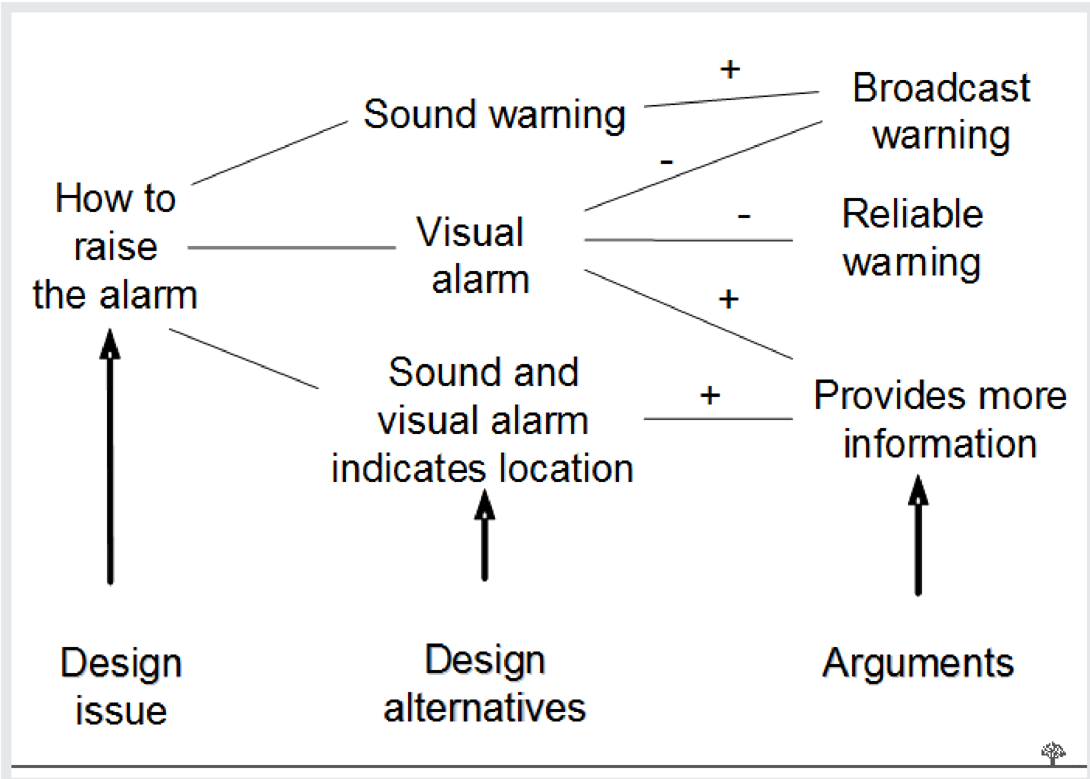



Figure 13.13: Design Rationale Diagram showing the gIBIS-Requirements Engineering variant.



The design issues illustrated in Figure 13.13 comes from a case study of requirements analysis for a safety-critical fire-management system on container ships. Heat sensors detected fire, then the issue was how to notify the crew. The diagram shows three design alternatives, linked to trade-off criteria. The top option proposes an audio alarm which does broadcast the alarm throughout the ship but does not give vital information on the location of the fire. The middle option of a visual alarm on a panel display is reliable but more localised, while the third option of using visual diagrams of the ship and highlighting the location of the fire provides better information. Design rationale diagrams allow users to see the trade-offs between different requirements as design options; however, QOC has been difficult to introduce into new communities of practice (MacLean & McKerlie, 1995), and similar problems have been encountered with the gIBIS (Conklin & Begeman, 1988) version of design rationale (Buckingham Shum, 1996; Sutcliffe & Ryan, 1997).

## 13.4 DESIGN PROCESS IN REQUIREMENTS ENGINEERING AND HUMAN-COMPUTER INTERACTION

One of the key differences between Requirements Engineering and HCI lies in process. Requirements Engineering is a systematic engineering discipline, so techniques and a systematic process are favoured. While Requirements Engineering has no standard methods, such as RUP in SE, Requirements Engineering nevertheless follows a process and applies a set of techniques to elicit, analyse, specify and validate requirements, for example the Volere method (Robertson & Robertson, 2002). In Software Engineering the nature of process guidance can vary from the substantial, as in RUP, to lightweight agile methods (Beck, 1999), which are similar to scenario-based design.

In HCI, development proceeds by user-centred design where iterative cycles of design exploration and user evaluation gradually refine the design to converge on a solution acceptable to users' needs. No formal procedures for checking or testing the design are adopted. In contrast, Requirements Engineering places more emphasis on automated checking of specifications, often referred to as model checking. This involves automated checking of software specifications to ensure that the program code will eventually operate correctly and reliably. One of the most influential formal approaches to Requirements Engineering is the KAOS method and associated tools (Van Lamsweerde, 2009). KAOS follows the Requirements Engineering mainstream with a goal-oriented approach but takes specification further as goals are refined to lower-level detail expressing states that the system should achieve, maintain or avoid. Model checkers then validate whether the goals can be achieved by the specified procedures. KAOS supports Requirements Engineering with tools, so constraints, assumptions and barriers to goals, called obstacles, can be assessed; for instance, what assumptions are necessary about <resources, processes, conditions> for a particular goal to be achieved? This just gives the gist of what KAOS and its support tools GRAIL can do; in reality it is more complex, with a temporal logic that also enables reasoning about when goals can be achieved. However, KAOS and all formal Software Engineering approaches do assume that considerable detail is known about the specification and the system environment. This is a key point of divergence with HCI, which conceptualises development as a more dynamic process where knowledge is partial.HCI used to have an interest in formal analysis but this has waned over the years, although it is still active (Thimbleby, 2007) with similar interests in model checking.

## 13.5 DESIGN ADVICE AND KNOWLEDGE REUSE

Both HCI and Requirements Engineering reuse knowledge, although in different ways. In Requirements Engineering, product-line requirements are the main approach to reuse, in which requirements are associated with a set of related designs in a particular application domain. Product lines are variations on a theme, common in engineering sectors; for example, automotive industry software for engine control and braking systems. Requirements can be categorised as common core requirements shared by many applications and variation points: requirements which change between different design versions. Requirements Engineering has adopted methods from the wider field of software reuse which creates hierarchy diagrams of features, showing points in the feature decomposition where tailoring may occur. In Requirements Engineering, processes have been researched for managing product-line requirements to plan releases and to prioritise variations and versions. (Finkelstein et al., 2008).

Requirements reuse has received less attention, although some generalised requirements have been proposed in studies of aero engine controllers (Lam et al., 1997); and, at a more abstract level, a library of reusable, generic requirements was associated with abstract problem models (Sutcliffe,2002a). These abstract models are distant relatives of more detailed and domain-specific product lines, so reuse is intended to provide tools for thought during the requirement process rather than specifications which can be customised. In Requirements Engineering the most abstract reuse model is Problem Frames (Jackson, 2001), which describes high-level requirements in terms of dependencies between software processes and their external environment. Four problem frames describe monitoring and controlling devices (required behaviour), responding to external commands (commanded behaviour), editing and updating states (workpieces) and general transformations. Abstract problem models provide more detail with 11 families of problems including transactions, hiring, monitoring control, logistics and tracking.

HCI places considerable emphasis on reuse of knowledge. Copious quantities of guidelines have been produced (ISO1997, 1998) with principles and heuristics as a more general expression of design advice (Benyon & Macaulay, 2002). The task-artefact cycle, illustrated in Figure 13.14 (Carroll, 2000), promotes reuse via artefacts which are design examples that can be organised in families similar to product lines (Sutcliffe, 2000b), although artefacts tend to be motivated by interaction design concerns rather than by functionality, as in product-line requirements. Artefacts are associated with claims which encapsulate design trade-offs as upsides and downsides in a similar manner to design rationale. Claims are also associated with scenarios of use to anchor the design advice in a realistic setting.

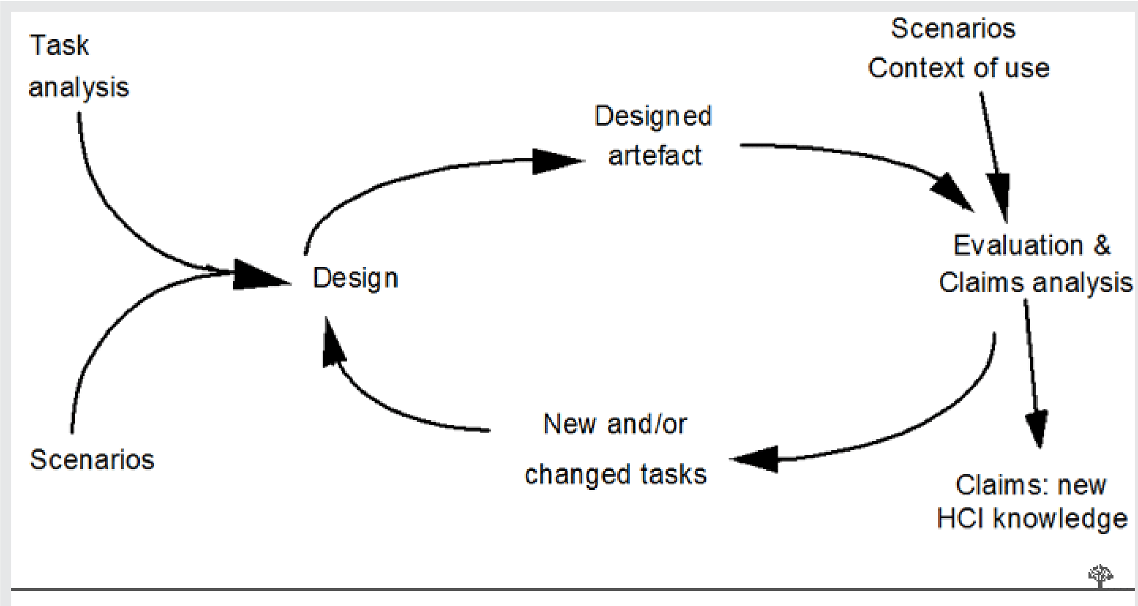



Figure 13.14: Task-artefact cycle for creating reusable knowledge in HCI.



The upsides and downsides in claims present usability arguments as trade-offs in psychologically based design rationale. Claims are generated by implementing a design, evaluating the usability of the implemented artefact, then extracting the claim and iteratively improving the design. The artefact becomes an example that instantiates the claim, while the claim and links to theory provide justification about why the claim should deliver usability, with supplementary information on dependencies and design trade-offs. See Figure 13.15.

| | |
|---|---|
| Claim ID: | Colour-coded Telegraphic Display |
| Author: | Singley, M.K.; Carroll, J.M. |
| Artifact: | MoleHill tutor - Goalposter tool |
| Description: | A colour-coded telegraphic display of goals |
| Upside: | Provides persistent feedback on the correctness of actions, as well as access to further information |
| Downside: | Learners must learn the display's feature-language and controls |
| Scenario: | The presentation of individual goals in the window is telegraphic, several words at most. However, the learner can expand any of the telegraphic goals (through a menu selection) to display a fuller explanation of why the goal is worthwhile pursuing or not. Thus the system provides both shorthand feedback on correctness and access to further help. |

FIGURE 13.15: Example of a claim (after Carroll et al., 1992).

Claims offer more targeted delivery of HCI knowledge than do guidelines, by virtue of the task-artefact cycle that provides a context. Claims have a domain-specific anchor in the artefact context; however, they can be linked to generic problem models to give a wider context for reuse (Sutcliffe & Carroll, 1998; Sutcliffe, 2002a); for example, human factors problems are associated with monitoring tasks or transaction processing for hiring/loans, sales, and logistics/distribution, etc. Furthermore, claims can be associated with interaction patterns to relate HCI knowledge to Software Engineering use cases.

**-- Purchase Process --**

**Problem**

Users want to purchase an already selected product

**Solution**

Present users with the purchase steps

[Example from www.bn.com]

**Use when**

The site allows purchasing of goods, typically an E-commerce Site but it can also be a site that happens to sell products as well such as a Museum Site. A purchase can also be part of larger tasks such as a Booking.

**How**

In order to purchase the products in the cart they need to select the checkout action. The checkout is a five step Purchase Process with the following tasks:

1. Identify the client
2. Select shipping address and special options
3. Select payment method
4. See overview of the entire order
5. Confirm and place order
6. Receive confirmation by email

The users can abort the checkout procedure at any step. When users retry the checkout later, they start again at the first task. Consider a Wizard to guide the user through these tasks while minimizing the number of web pages used. However, a wizard is not always needed for just a purchase. Often sites ask for details that are not strictly necessary to process the order. In many cases, all of the order information may easily fit on one page and hence eliminating the need for a wizard.

**Minimize navigation and non-relevant page elements**

Since purchasing is a task that requires quite some focus, the standard page layout during the purchase process has to be simplified. Sub-navigation and contextual elements should not be shown. All distracting elements should be removed.

**User Login**

Many sites require users to Login as the first step of the process. While this is convenient for returning customers because all their personal data can be re-used, it is not very nice for new users. New customers should be allowed to purchase items without creating an account. At the end of a purchase, users can be asked to Registration. Registration can then be made very simple because all the basic data has already been captured during the purchase process, only the username and password still needs to be selected.

**Confirmation by email**

It is important to “give” the users something that is easily accessible after the browser has been closed. An email with the information about the purchase is like a “receipt” for users. It should contain an order number, list of items in the order, all amount, shipping address, payment information, date of placing order. It should also contain help for users how to track they order, cancel it, or request assistance.

**Why**

First time customers or infrequent customers are best helped with a Wizard that allows the user to complete the purchase in small steps. Returning customers usually use the same shipping address and same credit-card. Therefore the process can be more efficiently done in only one overview screen with a “purchase” button.

Figure 13.16: Example of an HCI pattern for design of e-commerce purchase interaction.

This view of claims is similar to patterns which are shared by both disciplines. Patterns record reusable design advice with an example of its use, a motivating scenario, description of the contents for application,and forces: essentially trade-offs of advantages and disadvantages. In HCI, patterns tend to focus on user interface problems and designs (Tidwell, 2005), while Software Engineering patterns range from low-level software solutions (Gamma et al., 1995) to higher-level processes and designs (Coplein, 1996). Patterns recommend that design advice is presented in the context of a motivating problem, and with an example of its application (Borchers, 2001). Although patterns do have a clause that indicates the range of problems the design advice can be applied to, this scoping is *ad hoc*. Advocates of patterns propose relationships between individual patterns constructed into a hypertext-like pattern network or language (Alexander et al., 1977) to set the context. Unfortunately, pattern languages are also *ad hoc* and tend to be incomplete.

An example of an HCI pattern from the van Welie collection (www.welie.com) is illustrated in abbreviated form in Figure 13.16.

The pattern describes the problem context that it applies to and is then gives a solution with examples. Links are given to related patterns and some pattern formats add Forces to give arguments about the advantages and disadvantages of using the pattern.

## 13.6 CONCLUSIONS AND FUTURE DIRECTIONS

HCI and RE hold many common views about the design process. Both disciplines advocate iterative design cycles with testing and evaluation, although the degree of user-centredness varies. One topic neglected by both disciplines is deciding the automation boundary. This has been researched in human factors where methods and heuristics for functional allocation decide which process should be fully automated, manually operated or shared human-computer cooperation (Wright et al., 2000; Sutcliffe, 2002c). However, in Requirements Engineering this issue is rarely addressed and mainstream HCI assumes that a user interface boundary emerges from the process of design exploration.

While HCI has become less method-oriented with time, Requirements Engineering has maintained a more systematic engineering approach. Methods, guidelines, principles and models are rarely used explicitly by expert designers (Guindon & Curtis, 1988). Scenario-based design is the closest that HCI comes to a systematic method, although the approach may be regarded as tools for thought rather than step-by-step guidance. Scenarios, as Carroll suggests, support the design process at runtime as probes to test assumptions and stimulate creation. Scenarios can stimulate thought, but knowledge can only be reused effectively in a generalised form as claims, principles and guidelines.

HCI has expanded its focus from goal-oriented office work-style applications to user experience in entertainment and product design (Hassenzahl, 2010; Sutcliffe, 2009), with the role of emotion and feelings being recognised as important requirements for successful interactive systems (Norman, 2004). Requirements Engineering, in contrast, has barely acknowledged that goal-oriented applications exist apart from the occasional treatment of emotional requirements (Callele et al., 2006) and values (Thew et al., 2008). Requirements Engineering needs to expand its endeavour to encompass user-experience requirements and value-based design (Cockton, 2009).

On the other hand, HCI can benefit from Requirements Engineering by rediscovering past research into systematic specification where these concerns are paramount in safety-critical applications. Techniques for obstacle analysis (Van Lamsweerde, 2009) and reasoning about assumptions and user preferences (Jureta et al., 2008) could be usefully incorporated into HCI.

Finally, both disciplines need to recognise the changing nature of software as applicationsbecome more intelligent, self-aware and adaptable. In Requirements Engineering, requirements for intelligent applications are still research agendas (Sawyer et al., 2010), while in HCI intelligent user interfaces are present in recommenders and attentional user interfaces as well as occupying the IUI conference series in their own right. However, requirements and human-centric design of intelligent and socially networked software are future challenges for both disciplines.

## 13.7 WHERE TO LEARN MORE

### 13.7.1 Textbooks

The most comprehensive text book is Van Lamsweerde (2009), *Requirements engineering: From system goals to UML models to software specifications*. Chichester: Wiley.

Lamsweerde, Axel van (2009): *Requirements Engineering: From System Goals to UML Models to Software Specifications*. Wiley

### 13.7.2 Journals and Conference Series

The main journal of the discipline is*Requirements Engineering* (Springer). *Proceedings* of the IEEE International Conference on Requirements Engineering are available from 2002 onwards; before 2002 there were two separate conferences: the IEEE Symposium of Requirements Engineering, and the International Conference on Requirements Engineering.

**Requirements Engineering**

2009 2007 2004 1998

**RE - IEEE International Conference on Requirements Engineering**

2008 2007 2006 2005 2004 2003 2002 2001
2000 1999 1998 1997 1996 1995

See also *IEEE Software, Information and Software Technology, Journal of Information Technology*.

### IEEE Software Magazine

2002

### Information and Software Technology

2010 2007 1999 1997 1996

### Journal of Information Technology

2006

## 13.7.3 Internet

The Requirements Engineering Specialist Group (of the British Computer Society, www.resg.org.uk/) provides a comprehensive listing of RE resources, tools and a bibliography;the group publishes a *Newsletter* listing events with reviews and informal articles.

## 13.7.4 Related Fields

The journals *Human Computer Interaction*, *Software Engineering*, *Proceedings* of the International Conference on Software Engineering (ICSE) and the *IEEE Transactions on Software Engineering* publish RE-related articles. See *Proceedings* of the Participatory Design Conferences for a Scandiavian view of the RE process.

### IEEE Transactions on Software Engineering

1992 1981

### Human Computer Interaction

2011 2010 2007 2006

## PDC - International Conference on Participatory Design

2012 2012 2006 2004

## 13.8 REFERENCES


Alexander, Christopher, Ishikawa, Sara and Silverstein, Murray (1977): *A Pattern Language*. Oxford University Press

Beck, Kent (1999): *Extreme Programming Explained: Embrace Change*. Reading, MA, Addison-Wesley Publishing

Bellotti, Victoria (1988): Implications of Current Design Practice for the Use of HCI Techniques. In: Jones, Dylan M. and Winder, R. (eds.) *Proceedings of the Fourth Conference of the British Computer Society Human Computer Interaction Specialist Group - People and Computers IV* August 5-9, 1988, University of Manchester, UK. pp. 13-34

Benyon, David and Macaulay, Catriona (2002): Scenarios and the HCI-SE design problem. In *Interacting with Computers*, 14 (4) pp. 397-405

Boehm, Barry W. (1981a): *Software Engineering Economics*. Prentice Hall

Boehm, Barry W (1981b): Software Engineering Economics. In *IEEE Transactions on Software Engineering*, 16 (1) pp. 4-21

Boehm, Barry, Bose, Prasanta, Horowitz, Ellis and Lee, Ming-June (1994): Software requirements as negotiated win conditions. In *Proceedings of IEEE International Conference on Requirements Engineering*, pp. 74-83

Borchers, Jan O. (2001): *A Pattern Approach to Interaction Design*. John Wiley and Sons

Borchers, Jan O. (2000): A Pattern Approach to Interaction Design. In: *Proceedings of DIS00: Designing Interactive Systems: Processes, Practices, Methods, & Techniques* 2000. pp. 369-378

Bubenko, Janis A. and Kirikova, Marite (1994). Enterprise Modelling: Improving the Quality of Requirements Specifications.

Callele, David, Neufeld, Eric and Schneider, Kevin (2006): Emotional Requirements in Video Games. In: *14th IEEE International Conference on Requirements Engineering RE 2006* 11-15 September, 2006, Minneapolis/St.Paul, Minnesota, USA. pp. 292-295

*Carroll, John M. (2002): Making use is more than a matter of task analysis. In Interacting with Computers, 14 (5) pp. 619-627*

*Carroll, John M. (2000): Making Use: Scenario-Based Design of Human-Computer Interactions. MIT Press*

*Carroll, John M. (ed.) (1995): Scenario-Based Design: Envisioning Work and Technology in System Development. John Wiley and Sons*

*Carroll, John M., Singley, Mark K. and Rosson, Mary Beth (1992): Integrating Theory Development with Design Evaluation. In Behaviour and Information Technology, 11 (5) pp. 247-255*

*Carroll, John M., Alpert, Sherman R., Karat, John, Deusen, Mary S. Van and Rosson, Mary Beth (1994): Raison d'Etre: capturing design history and rationale in mutimedia narratives. In: Plaisant, Catherine (ed.) Conference on Human Factors in Computing Systems, CHI 1994, Boston, Massachusetts, USA, April 24-28, 1994, Conference Companion 1994. p. 213*

*Checkland, Peter (1981): Systems Thinking, Systems Practice: Includes a 30-Year Retrospective. Wiley*

*Cockburn, Alistair (2000): Writing Effective Use Cases. Addison-Wesley Professional*

*Cockton, Gilbert (2009): Getting there: six meta-principles and interaction design. In: Proceedings of ACM CHI 2009 Conference on Human Factors in Computing Systems 2009. pp. 2223-2232*

*Conklin, Jeff and Begeman, Michael L. (1988): gIBIS: A Hypertext Tool for Exploratory Policy Discussion. In: Greif, Irene (ed.) Proceedings of the 1988 ACM conference on Computer-supported cooperative work September 26 - 28, 1988, Portland, Oregon, United States. pp. 140-152*

*Conklin, Jeff and Begeman, Michael L. (1988a): gIBIS: A Hypertext Tool for Exploratory Policy Discussion. In ACM Transactions on Information Systems, 6 (4) pp. 303-331*

*Dardenne, Anne, Lamsweerde, Axel van and Fickas, Stephen (1993): Goal-directed requirements acquisition. In Science of Computer Programming, 20 (1) pp. 3-50*

*Davenport, Thomas H. (1992): Process Innovation: Reengineering Work Through Information Technology. Harvard Business Press*

*Dearden, Andy, Harrison, Michael and Wright, Peter C. (2000): Allocation of Function: Scenarios, Context and the Economics of Effort. In International Journal of Human-Computer Studies, 52 (2) pp. 289-318*

*Diaper, Dan (2001): Task analysis for knowledge descriptions (TAKD): a requiem for a method. In Behaviour and Information Technology, 20 (3) pp. 199-212*

*Diaper, Dan and Johnson, Peter (1989): Task analysis for knowledge descriptions: Theory and application in training. In: Long, John and Whitefield, A. (eds.). "Cognitive Ergonomics and Human-Computer Interaction (Cambridge Series on Human-Computer Interaction)". Cambridge University Press*

*Djajadiningrat, J. P., Gaver, William W. and Fres, J. W. (2000): Interaction Relabelling and Extreme Characters: Methods for Exploring Aesthetic Interactions. In: Proceedings of DIS00: Designing Interactive Systems: Processes, Practices, Methods, & Techniques 2000. pp. 66-71*

*Eason, K. D., Harker, S. D. P. and Olphert, C. W. (1996): Representing socio-technical systems options in the development of new forms of work organisation. In European Journal of Work and Organisational Psychology,*

*Eden, Colin (1988): Cognitive mapping. In European Journal of Operational Research, 36 (1) pp. 1-13*

*Fenton, Norman E. (1995): The role of measurement in software safety assessment. In: 12th Annual CSR/ENCRESS Conference on Software Safety Case 13 Sept, 1995, Bruges, Belgium.*

*Finkelstein, Anthony, Harman, Mark, Mansouri, S. Afshin, Ren, Jian and Zhang, Yuanyuan (2008): "Fairness Analysis" in Requirements Assignments. In: Proceedings of the 2008 16th IEEE International Requirements Engineering Conference 2008. pp. 115-124*

*Gamma, Erich, Helm, Richard, Johnson, Ralph and Vlissides, John (1995): Design Patterns: Elements of reusable object-oriented software. Addison-Wesley Publishing*

*Gause, Donald C. and Weinberg, Gerald M. (1989): Exploring Requirements: Quality Before Design. Dorset House Publishing Company, Incorporated*

*Goguen, Joseph A. and Linde, Charlotte (1993): Techniques for requirements elicitation. In 1993 Proceedings of the IEEE International Symposium on Requirements Engineering, pp. 152-164*

*Gould, John D. (1987): How to Design Usable Systems. In: Bullinger, Hans-Jorg and Shackel, Brian (eds.) INTERACT 87 - 2nd IFIP International Conference on Human-Computer Interaction September 1-4, 1987, Stuttgart, Germany. pp. xxxv-xli*

*Graham, Ian M. (1996): Task scripts, use cases and scenarios in object oriented analysis. In Object Oriented Systems, 3 pp. 123-142*

*Graham, Ian (): Task scripts, use cases and scenarios in objectoriented analysis. In: Object-Oriented Systems .*

*Guindon, Raymonde and Curtis, Bill (1988): Control of Cognitive Processes During Software Design: What Tools Are Needed?. In: Soloway, Elliot, Frye, Douglas and Sheppard, Sylvia B. (eds.) Proceedings of the ACM CHI 88 Human Factors in Computing Systems Conference June 15-19, 1988, Washington, DC, USA. pp. 263-268*

*Hassenzahl, Marc (2010): Experience Design: Technology for All the Right Reasons. Morgan and Claypool Publishers*

*Hauser, John R. and Clausing, Don (1988): The House of Quality. In Harvard Business Review, 66 (3) pp. 63-73*

*Holland, Christopher P. (1995): Cooperative supply chain management: the impact of interorganizational information systems. In Journal of Strategic Information Systems, 4 (2) pp. 117-133*

*Hughes, John A., O'Brien, Jon, Rodden, Tom, Rouncefield, Mark and Sommerville, Ian (1995): Presenting ethnography in the requirements process. In: Second IEEE International Symposium on Requirements Engineering 1995 March 27 - 29, 1995, York, England. pp. 27-39*

*Jackson, Michael (2000): Problem Frames: Analysing & Structuring Software Development Problems. Addison-Wesley Professional*

*Jackson, Michael (2001): Problem Frames: Analysing & Structuring Software Development Problems. Addison-Wesley Professional*

*Jackson, Michael and Zave, Pamela (1993): Domain Descriptions. In RE, pp. 56-64*

*Jacobson, Ivar, Christerson, Magnus, Jonsson, Patrik and Övergaard, Gunnar (1992): Object-oriented software engineering: A use-case driven approach. Reading, MA, Addison-Wesley Publishing*

*Johnson, Peter, Johnson, Hilary, Waddington, Ray and Shouls, Alan (1988): Task-Related Knowledge Structures: Analysis, Modelling and Application. In: Jones, Dylan M. and Winder, R. (eds.) Proceedings of the Fourth Conference of the British Computer Society Human Computer Interaction Specialist Group - People and Computers IV August 5-9, 1988, University of Manchester, UK. pp. 35-62*

*Johnson, W. Lewis, Feather, Martin S. and Harris, David R. (1992): Representation and Presentation of Requirements Knowledge. In IEEE Trans. Software Eng., 18 (10) pp. 853-869*

*Jureta, Ivan, Mylopoulos, John and Faulkner, Stéphane (2008): Revisiting the Core Ontology and Problem in Requirements Engineering. In: 16th IEEE International Requirements Engineering Conference RE 2008 8-12 September, 2008, Barcelona, Catalunya, Spain. pp. 71-80*

*Kaindl, Hermann (1995): An integration of scenarios with their purposes in task modeling. In: Proceedings of the 1st conference on Designing interactive systems processes, practices, methods, and techniques 1995. pp. 227-235*

*Kethers, Stefanie and Jacobs, Stephan (1994): Improving Communication and Decision Making within Quality Function Deployment. In: First International Conference on Concurrent Engineering, Research, and Application 1994.*

*Kuuti, Kari (1995): Workprocess: Scenarios as a preliminary vocabulary. In: Carroll, John M. (ed.). "Scenario-Based Design: Envisioning Work and Technology in System Development". John Wiley and Sonspp. 19-36*

*Kyng, Morten (1995): Creating contexts for design. In: Carroll, John M. (ed.). "Scenario-Based Design: Envisioning Work and Technology in System Development". John Wiley and Sonspp. 85-108*

*Lam, W., McDermid, John A. and Vickers, Andrew (1997b): Ten Steps Towards Systematic Requirements Reuse. In Requir. Eng., 2 (2) pp. 102-113*

*Lam, W., McDermid, John A. and Vickers, Andrew (1997): Ten Steps Towards Systematic Requirements Reuse. In: 3rd IEEE International Symposium on Requirements Engineering RE97 January 5-8, 1997, Annapolis, MD, USA. pp. 6-15*

*Lamsweerde, Axel van (2009): Requirements Engineering: From System Goals to UML Models to Software Specifications. Wiley*

*Lamsweerde, Axel van and Letier, Emmanuel (2000): Handling Obstacles in Goal-Oriented Requirements Engineering. In IEEE Trans. Software Eng., 26 (10) pp. 978-1005*

*Lamsweerde, Axel van, Darimont, R. and Massonet, P. (1995): Goal-directed elaboration of requirements for a meeting scheduler: problems and lessons learnt. In RE, pp. 194-203*

*Luff, Paul, Jirotka, Marina, Heath, Christian C. and Greatbatch, David (1993): Tasks and social interaction: the relevance of naturalistic analyses of conduct for requirements engineering. In: Proceedings of IEEE International Symposium on Requirements Engineering 1993 1993, San Diego, CA , USA . pp. 187-190*

*Macaulay, L. (1993): Requirements capture as a cooperative activity. In: Proceedings of IEEE International Symposium on Requirements Engineering 1993 1993, San Diego, CA , USA . pp. 174-181*

*MacLean, Allan, Young, Richard M., Bellotti, Victoria and Moran, Thomas P. (1991): Questions, Options, and Criteria: Elements of Design Space Analysis. In Human-Computer Interaction, 6 (3) pp. 201-250*

*Maiden, N. A. M. and Ncube, C. (1998): Acquiring requirements for Commercial Off-The-Shelf package selection. In IEEE Software, 15 (2) pp. 46-56*

*Maiden, N. A. M. and Rugg, G. (1994): Knowledge acquisition techniques for requirements engineering. In: Proceedings of the Workshop on Requirements Elicitation for System Specification 12-14 July, 1994, Keele, University of Keele.*

*Maiden, N. A. M. and Sutcliffe, Alistair G. (1994): Requirements critiquing using domain abstractions. In: Requirements Engineering 1994 - Proceedings of the First International Conference on Requirements Engineering 1994, Colorado Springs, CO , USA. pp. 184-193*

*Monk, Andrew, Wright, Peter, Haber, Jeanne and Davenport, Lora (1993): Improving Your Human-Computer Interface: A Practical Technique (Bcs Practitioner). Prentice Hall*

*Mylopoulos, John, Chung, Lawrence and Nixon, Brian (1992): Representing and using non-functional requirements: a process-oriented approach. In IEEE Transactions on Software Engineering, 18 (6) pp. 483-497*

*Mylopoulos, John, Chung, Lawrence and Yu, Eric S. K. (1999): From Object-Oriented to Goal-Oriented Requirements Analysis. In Communications of the ACM, 42 (1) pp. 31-37*

*Norman, Donald A. (2004): Emotional Design: Why We Love (Or Hate) Everyday Things. Basic Books*

*Nuseibeh, Bashar and Easterbrook, Steve M. (2000): Requirements engineering: a roadmap. In: ICSE - Future of SE Track 2000 2000. pp. 35-46*

*Pohl, Klaus (1996): Process-Centered Requirements Engineering (Advanced Software Development Series). Research Studies Press*

*Porter, Michael E. (1980): Competitive Strategy: Techniques for Analyzing Industries and Competitors. Free Press*

*Potts, Colin (1999): ScenIC: A Strategy for Inquiry-Driven Requirements Determination. In: 4th IEEE International Symposium on Requirements Engineering RE 99 7-11 June, 1999, Limerick, Ireland. pp. 58-65*

*Potts, Colin, Takahashi, K. and Anton, A. I. (1994): Inquiry-based requirements analysis. In IEEE Software, 11 (2) pp. 21-32*

*Potts, Colin, Takahashi, Kenji, Smith, Jeffrey D. and Ota, Kenji (1995): An evaluation of inquiry-based requirements analysis for an Internet service. In: 1995. pp. 172-180*

*Robertson, Suzanne and Robertson, James (1999): Mastering the Requirements Process. Addison-Wesley Professional*

*Rockart, J. F. and Short, J. E. (1991): The networked organisation and the management of interdependence. In: Morton, Michael S. Scott (ed.). "The Corporation of the 1990s: Information Technology and Organizational Transformation". Oxford University Press, USA*

*Rodden, Tom (1999): Human factors in requirements engineering: - A survey of human sciences literature relevant to the improvement of dependable systems development processes. In Interacting with Computers, 11 (6) pp. 665-698*

*Rodriguez, F. Gamboa and Scapin, D. L. (1997): Editing MAD* task description for specifying interfaces, at both semantic and presentation level. In: Design, specification and verification of interactive systems 1997 1997.*

*Rolland, C., Achour, C. Ben, Cauvet, C., Ralyté, J., Sutcliffe, Alistair G., Maiden, N., Jarke, M., Haumer, P., Pohl, Klaus, Dubois, E. and Heymans, P. (1998): A proposal for a scenario classification framework. In Requirements Engineering, 3 (1) pp. 23-47*

*Sawyer, Peter, Bencomo, Nelly, Whittle, Jon, Letier, Emmanuel and Finkelstein, Anthony (2010): Requirements-Aware Systems: A Research Agenda for RE for Self-adaptive Systems. In: RE 2010, 18th IEEE International Requirements Engineering Conference, Sydney, New South Wales, Australia, September 27 - October 1, 2010 2010. pp. 95-103*

*Shum, S. Buckingham (1996): Analyzing the usability of a design rationale notation. In: Moran, Thomas P. and Carroll, John M. (eds.). "Design Rationale: Concepts, Techniques, and Use (Computers, Cognition, and Work Series)". CRC Press*

*Stapleton, Jennifer and Constable, Peter (1997): DSDM: Dynamic Systems Development Method: The Method in Practice. Addison-Wesley Professional*

*Sutcliffe, Alistair G. (2002c): The Domain Theory: Patterns for Knowledge and Software Reuse. CRC Press*

*Sutcliffe, Alistair G. (1997): Task-Related Information Analysis. In International Journal of Human-Computer Studies, 47 (2) pp. 223-257*

*Sutcliffe, Alistair G. (1995): Requirements Rationales: Integrating Approaches to Requirement Analysis. In: Proceedings of DIS95: Designing Interactive Systems: Processes, Practices, Methods, & Techniques 1995. pp. 33-42*

*Sutcliffe, Alistair G. (2000): Bridging the communications gap: developing a Lingua Franca for software developers and users. In: Actes du XVIIIe Congres 2000 2000, Toulouse, France. pp. 13-32*

*Sutcliffe, Aiistair (2009): Designing for User Engagement: Aesthetic and Attractive User Interfaces. In Synthesis Lectures on HumanCentered Informatics, 2 (1) pp. 1-55*

*Sutcliffe, Alistair G. (2002): Modelling collaboration in loosely coupled inter-organisational relationships. In: Systems engineering for business process change 2002. pp. 347-366*

*Sutcliffe, Alistair G. (2000a): On the Effective Use and Reuse of HCI Knowledge. In ACM Transactions on Computer-Human Interaction, 7 (2) pp. 197-221*

*Sutcliffe, Alistair G. (2002a): User-Centred Requirements Engineering. Springer*

*Sutcliffe, Alistair G. and Carroll, John M. (1998): Generalizing Claims and Reuse of HCI Knowledge. In: Johnson, Hilary, Nigay, Laurence and Roast, C. R. (eds.) Proceedings of the Thirteenth Conference of the British Computer Society Human Computer Interaction Specialist Group - People and Computers XIII August 1-4, 1998, Sheffield, UK. pp. 159-176*

*Sutcliffe, Alistair G. and Ryan, Michele (1997): Assessing the usability and efficiency of Design Rationale. In: Howard, Steve, Hammond, Judy and Lindgaard, Gitte (eds.) Human-Computer Interaction, INTERACT 97, IFIP TC13 Interantional Conference on Human-Computer Interaction, 14th-18th July 1997, Sydney, Australia 1997. pp. 148-155*

*Sutcliffe, Alistair G., Fickas, Stephen and Sohlberg, McKay Moore (2005): Personal and Contextual Requirements Engineering. In: 13th IEEE International Conference on Requirements Engineering RE 2005 29 August - 2 September, 2005, Paris, France. pp. 19-30*

*Thew, Sarah, Sutcliffe, Alistair G., Bruijn, Oscar de, McNaught, John, Procter, Rob, Venters, Colin and Buchan, Iain (2008): Experience in e-Science Requirements Engineering. In: 16th IEEE International Requirements Engineering Conference RE 2008 8-12 September, 2008, Barcelona, Catalunya, Spain. pp. 277-282*

*Thew, Sarah, Sutcliffe, Alistair G., Procter, Rob, Bruijn, Oscar de, McNaught, John, Venters, Colin C. and Buchan, Iain (2009): Requirements Engineering for E-science: Experiences in Epidemiology. In IEEE Software, 26 (1) pp. 80-87*

*Thimbleby, Harold (2007): Press On: Principles of Interaction Programming. The MIT Press*

*Tidwell, Jenifer (2005): Designing Interfaces: Patterns for Effective Interaction Design. O'Reilly and Associates*

*Veer, Gerrit C Van Der, Lenting, Bert F and Bergevoet, Bas A J (1996): GTA : Groupware Task Analysis - Modeling Complexity. In Acta Psychologica, 91 (3) pp. 297-322*

*Vicente, Kim J. (2000): HCI in the Global Knowledge-Based Economy: Designing to Support Worker Adaptation. In ACM Transactions on Computer-Human Interaction, 7 (2) pp. 263-280*

*Viller, Stephen, Bowers, John and Rodden, Tom (1999): Human Factors in Requirements Engineering: A Survey of Human Sciences Literature Relevant to the Improvement of Dependable Systems Development Processes. In Interacting with Computers, 11 (6) pp. 665-698*

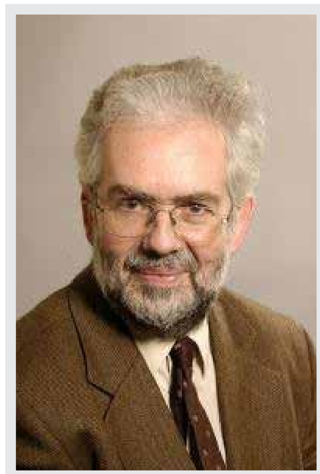
*Wright, Peter C., Dearden, Andy and Fields, Bob (2000): Function Allocation: A Perspective from Studies of Work Practice. In International Journal of Human-Computer Studies, 52 (2) pp. 335-355*

*Yu, E.S.K. (1993): Modeling organizations for information systems requirements engineering. In 1993 Proceedings of the IEEE International Symposium on Requirements Engineering, pp. 34-41*

*Zave, Pamela (1995): Classification of research efforts in requirements engineering. In: Second IEEE International Symposium on Requirements Engineering 1995 March 27 - 29, 1995, York, England. pp. 214-216*

## About the author

### Alistair G. Sutcliffe



*© Alistair G. Sutcliffe*

Has also published under the name of:
"Alistair Sutcliffe"

Alistair Sutcliffe (MA Cantab-Natural Sciences, PhD Wales) is Professor of Systems Engineering, and Director of the Centre for HCI Design, in the School of Informatics, University of Manchester, UK. Originally at ethologist, he has worked in the IT and finance industry, the civil service and City and Manchester Universities. His research spans software engineering, human computer interaction, cognitive and social science, with recent interests in scenario based design, methods for requirements engineering, analysis and modelling complex socio technical systems, visualisation and creative design . He is a leading authority on human factors in safety critical systems, requirements engineering and multimedia user interface design, has authored 6 books and 200+ publications on human computer interaction, requirements engineering, software and domain knowledge reuse. His recent books include: Multimedia and virtual reality: Designing multisensory user interfaces. Lawrence Erlbaum Associates (2003) User-Centred Requirements Engineering. Springer, (2002) and The Domain Theory: Patterns for knowledge and software reuse, Lawrence Erlbaum Associates (2002). He currently manages EPSRC projects ADVISES (E-science requirements analysis and visualisation) and ESRC/EPSRC Foresight project Developing Theory for Evolving Socio Technical Systems and was recently PI of EPSRC projects SIMP- Systems Integration for Major Projects, ISRE Immersive scenario based Requirements Engineering and CORK Corporate Knowledge Repository. He serves on the editorial boards of several journals in the software engineering and human computer interaction. He founded IFIP TC-13 Working Group 13.2 Methodology for User Centred Design, is editor of the ISO standard 14915, on Multimedia user interface design, co-chaired the ACM conference Designing Interactive Systems 2002 and is co-chair of IEEE Requirements Engineering conference in 2007. He was awarded the IFIP silver core in 1999, is a reviewer for EPSRC college, INRIA cognitive engineering projects 2002 & 2006, NSF Science of Design program and made recent plenary keynote presentations at IHM03, HCI05, and CAiSE06 conferences

## Your notes and thoughts on chapter 13

Record your notes and thoughts on this chapter. If you want to share these thoughts with others online, go to the bottom of the page at:

http://www.interaction-design.org/encyclopedia/requirements_engineering.html

**NOTES:**

---

---

# Chapter 14

# Context-Aware Computing

## Context-Awareness, Context-Aware User Interfaces, and Implicit Interaction

### *by Albrecht Schmidt.*

A tablet computer switching the orientation of the screen, maps orienting themselves with the user's current orientation and adapting the zoom level to the current speed, and switching on the backlight of the phone when used in the dark are examples of computers that are aware of their environment and their context of use. Less than 10 years ago, such functions were not common and existed only on prototype devices in research labs working on context-aware computing.

Figure 14.1: An iPad switching orientation of the screen is a good example of context-aware computing.



## 14.1 INTRODUCTION

When we aim to create applications, devices, and systems that are *easy to use*, it is essential to understand the *context of use*. With context-aware computing, we now have the means of considering the situation of use not only in the design process, but in real time while the device is in use. In Human-Computer Interaction (HCI), we traditionally aim to understand the user and the context of use and create designs that support the major anticipated use cases and situations of use. In Context-Aware Computing on the other hand, making use of context causes a fundamental change: We can support more than one context of use that are equally optimal. At runtime – when the user interacts with the application — the system can decide what the current context of use is and provide a user interface specifically optimized for this context. With context-awareness, the job of designing the user interface typically becomes more complex as the number of situations and contexts which the system will be used in usually increases. In contrast to traditional systems, we do not design for a single -or a limited set - of contexts of use; Instead, design for several contexts. The advantage of this approach is that we can provide optimized user interfaces for a range of contexts.

Let us assume the following example: You are asked to design a user interface for a wrist watch. In your research you find out that people will use it indoors and outdoors, they will use it in the dark as well as in sunlight, they will use it when they run to catch the train as well as when they sit in a lecture and are bored. As a good user interface designer, you end up with many ideas for an exciting user interface for each situation: For example, when the user is running to catch the train, the user interface should show the time highlighting the minutes and seconds in a very large font. On the other hand, when the user is attending a lecture the user interface should show the time in a very small font, and additionally provide a funny quote. In a traditional design process, you would realize – after creating your sketches and design briefs – that you have to decide which one of your ideas for a user interface you want to use. You would realize that supporting all the situations in a single design will not work. The typical result is a compromise – which often loses much of the edge of the ideas you initially came up with. However, if you take the approach of Context-Aware Computing, you could create a context-aware watch, where you combine all your situation-optimized designs in a single design. If you designed your watch so that it could recognize each of the situations that you had found in your initial research (e.g. running to catch the train, attending a lecture, etc), your watch could reconfigure itself based on the recognized context. Figure 14.2 shows a design sketch for a context-aware watch.

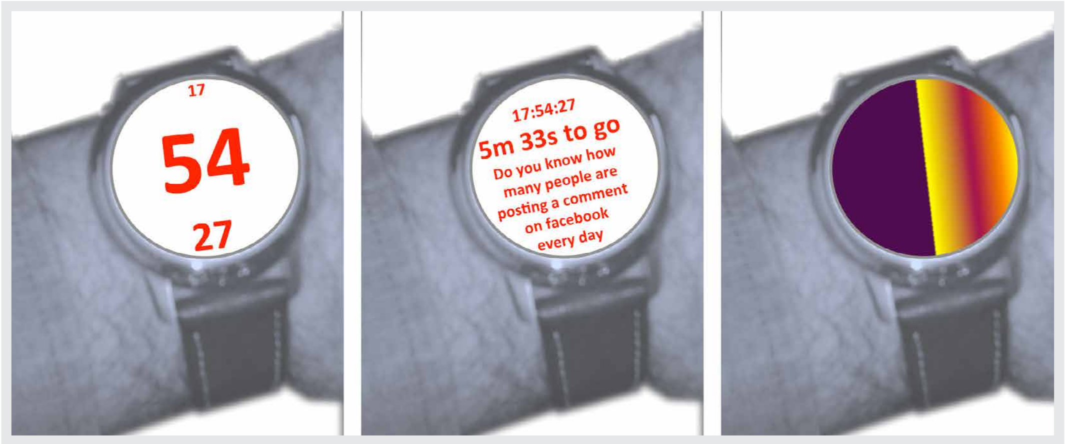



Figure 14.2: Design sketches that illustrate ideas for time visualizations in different contexts. Left: for users that run to catch the train; making it easy to see the minutes. Middle: a time visualization for boring lectures and meetings; shows a count down to the end, and some information to engage the user. Right: visualization that gives only a very coarse idea of the time, similar to the information you get from the sun, e.g. for hanging out with friends when time does not matter. With context-awareness, you could create a product that combines all three visualizations and choose the most appropriate one according to the recognized context.



The example shows the great advantage of context-aware computing systems as the freedom of design is increased, but at the same time systems become more complex and often more difficult to design and implement. In this chapter, we introduce the basics for creating context-aware applications and discuss how these insights may help design systems that are easier and more pleasant to use.

## 14.2 CONTEXT-AWARENESS

In the early days of computing, the context in which systems were used was strongly defined by the place in which computers were set up, see Figure 14.3. Personal computers were used in office environments or on factory floors. The context of use did not change much, and there was little variance in the situations surrounding the computer. Hence, there was no need to adapt to different environments. Many traditional methods in the discipline of Human-Computer Interaction (HCI), such as contextual inquiry or task analysis, have their origin in this period and are most easy to use in situations that do not constantly change. With the rise of mobile computers and ubiquitous computing, this changed. Users take computers with them and use them in many different situations, see Figure 14.4.

At the beginning of the mobile computing era, in the late 80s and 90s, the central theme was how to make mobility transparent for the user, automatically providing the same service everywhere. Here, transparent meant that the user did not have to care about changes in the environment and could rely on the same functionality independent of the environment.

In the early 90s, research into ubiquitous computing at Xerox PARC caused a shift in thinking. In addition to making functionality transparent, such as providing network connectivity throughout a campus without the user realizing the hand-over between different networks, researchers discovered the potential to exploit the context of use as a resource to which systems can be adapted. In his 1994 paper at the Workshop on Mobile Computing Systems and Applications (WMCSA), Bill Schilit introduces the concept of context-aware computing and describes it as follows:

> “ Such context-aware software adapts according to the location of use, the collection of nearby people, hosts, and accessible devices, as well as to changes to such things over time. A system with these capabilities can examine the computing environment and react to changes to the environment. ”
>
> *-- Schilit et al 1994*

The basic ideas is that mobile devices can provide different services in different contexts – where context is strongly related to the location of a device.

Much of the initial research of context-aware computing hence focused on *location-aware* systems. In this sense, the widely-used satellite navigation systems in cars today are context-aware systems. However, context is more than location, as we argue in (Schmidt et al 1999) and throughout this chapter.

*Courtesy of Library of Virginia. Copyright: pd (Public Domain (information that is common property and contains no original authorship)).*

*Courtesy of Library of Virginia. Copyright: pd (Public Domain (information that is common property and contains no original authorship)).*

*Courtesy of The Library of the London School of Economics and Political Science. Copyright: pd (Public Domain (information that is common property and contains no original authorship)).*

Figure 14.3 A-B-C: In the early days of computing, the context was defined by the computer as the computer was the actual workplace. Later, computers were set up in a specific location to help with a specific task, and hence the context was strongly defined by the location of the computer. Personal computers were used in office environments or on factory floors.

*Copyright status: Unknown (pending investigation). See section “Exceptions” in the copyright terms.*




FIGURE 14.4 A-B: Even in the early days of mobile computing, where notebook computers were considered mobile devices, users could choose the context in which to work. With the rise of mobile and handheld computers and ubiquitous computing, this changed even more radically. Users take computers with them and use them in many different situations in the real world. Next time you go for a walk, observe the multitude of mobile usage scenarios and you will be surprised what people do with their devices.

### 14.2.1 Example 1: SatNav as context-aware system

In a Satellite Navigation System (SatNav), the current location is the primary contextual parameter that is used to automatically adjust the visualization (e.g. map, arrows, directions...) to the user's current location. An example is shown in Figure 14.5. However, looking at current commercial systems, much more context information is used and much of visualization has been changed. In addition to the current GPS position, contextual parameters may include the time of day, light conditions, the traffic situation on the calculated route or the user's preferred places. Beyond the visualization and whether or not to switch on the backlight, the calculated route can be influenced by context, e.g. to avoid potentially busy streets at that time of day.


*Courtesy of Eirik Solheim. Copyright: CC-Att-SA-2 (Creative Commons Attribution-ShareAlike 2.0 Unported).*

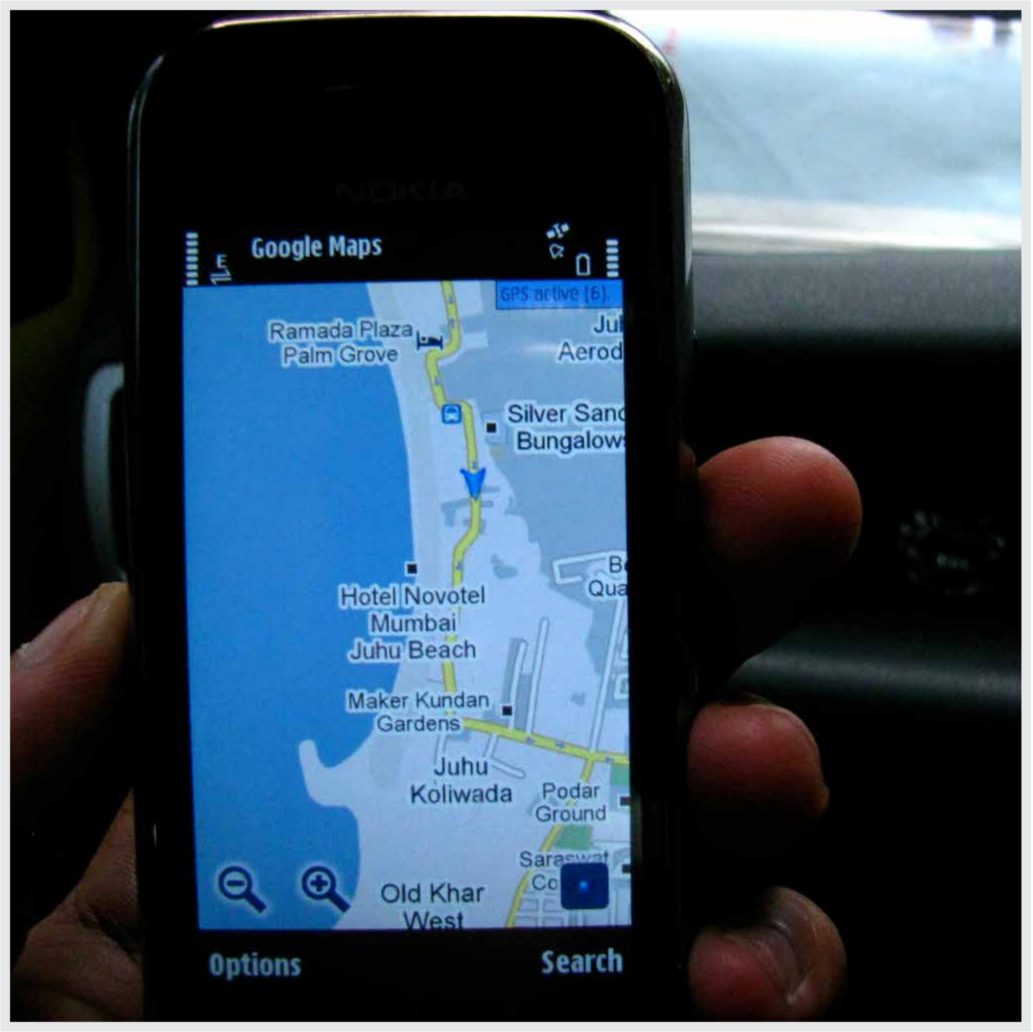





Figure 14.5 A-B: Navigation devices have become common and are widely used in cars and by pedestrians. Figure 4a shows a TomTom navigation application on a Nokia N95 device. Figure 4B shows Google Maps on another Nokia device. SatNavs are probably the most widespread context-aware computing systems.

### 14.2.2 Example 2: Automatic light as context-aware system

At house entrances and in hotel hallways automatic lights have become common. These systems can also be seen as simple context-aware systems. The contextual parameters taken into account are the current light conditions and if there is motion in the vicinity. The adaptation mechanism is fairly simple. If the situation detected is that it is dark and that there is someone moving, the light will be switched on. The light will then be on as long as the person moves, and after a period where no motion is detected, the light will switch off again. Similarly, the light will switch off if it is not dark anymore.

These simple examples outline the basic principle of a context-aware system. In Figure 14.6 a reference architecture for context-aware computing systems is shown. Sensors provide data about activities and events in the real world. Perception algorithms will make sense of these stimuli and classify the situations into context. Based on the observed context, actions of the system will be triggered.

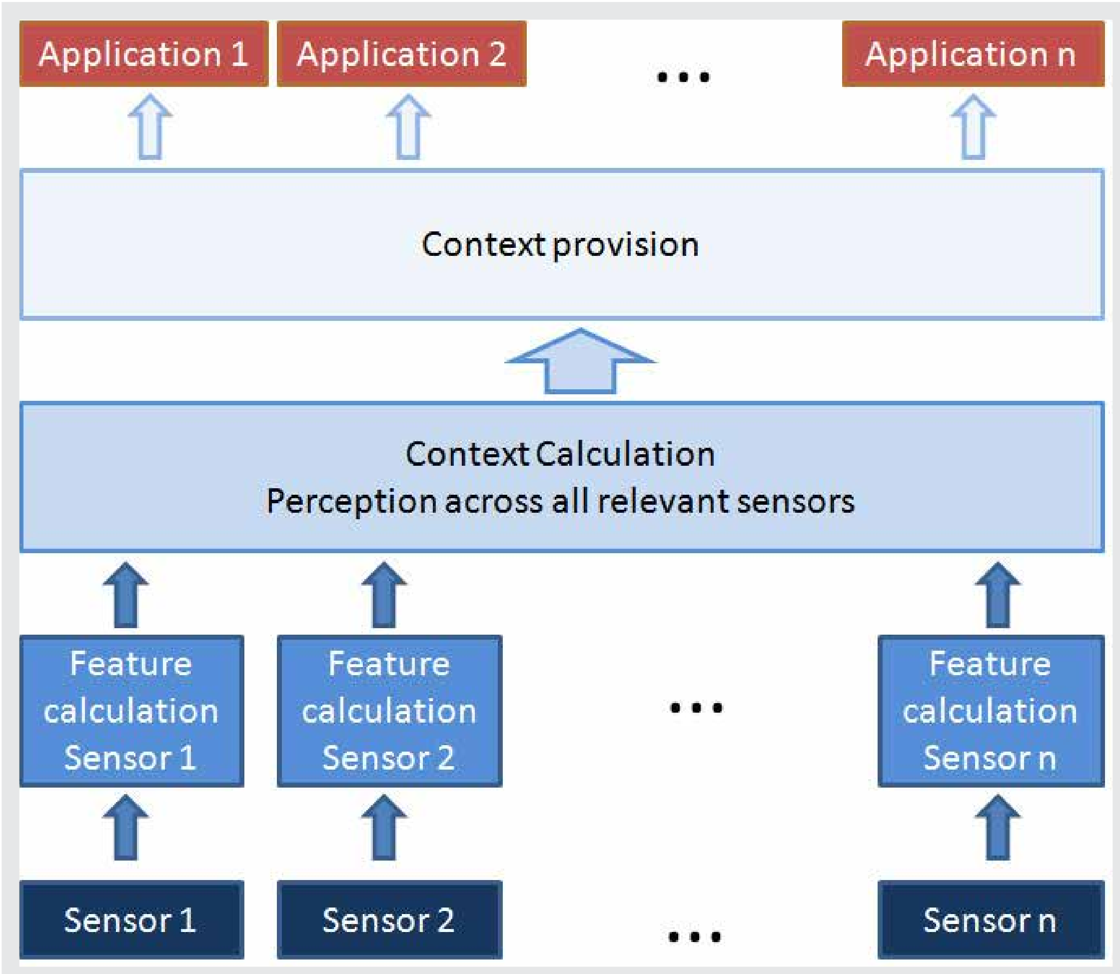



FIGURE 14.6: The drawing depicts a reference architecture for context-aware computing systems. It assumes the acquisition of data from sensors support to contextual behavior of multiple applications.



## 14.3 CONTEXT-AWARENESS AS ENABLER FOR UBIQUITOUS COMPUTING

The notion of context-awareness is closely related to the vision of ubiquitous computing, as introduced by Mark Weiser (see Figure 14.7) in his seminal paper in the journal Scientific American. As computers become a part of everyday life, it is essential that they are easy to use. This is highlighted by the following statement.

> “The most profound technologies are those that disappear. They weave themselves into the fabric of everyday life until they are indistinguishable from it.”
>
> -- *Weiser 1991*

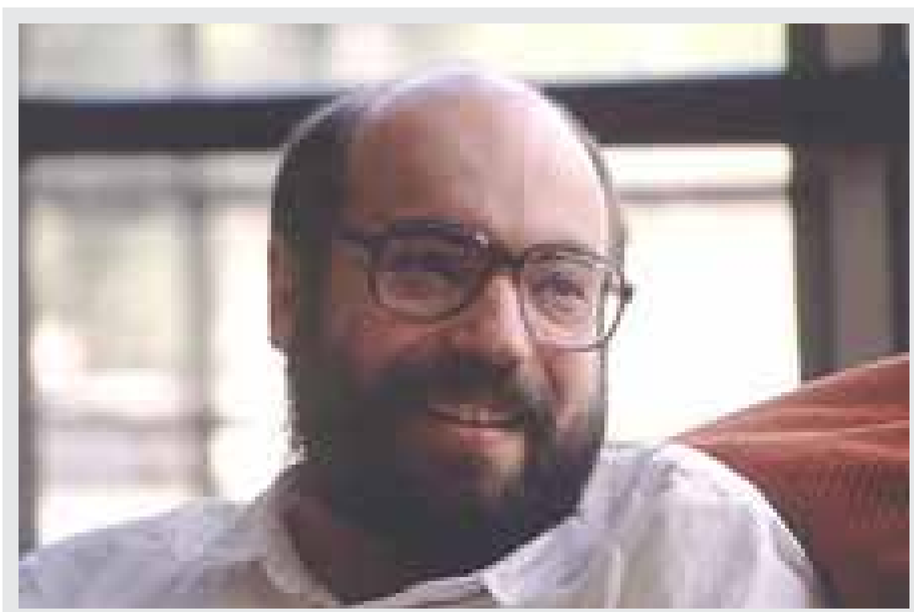


Figure 14.7: Mark Weiser stated the vision of ubiquitous computing. Context-awareness is an essential building block for realizing this vision.



Many people regard this level of integration of computing technologies as the ultimate goal for computers. In such a situation, technologies would not require much active attention by the user, and would be ready to use at a glance. If this is achieved, computers *disappear* – not in a technical sense, but from a psychological perspective.

> “ In essence, that only when things disappear in this way are we freed to use them without thinking.”
>
> -- *Weiser 1991*

To realize such ubiquitous computing systems with optimal usability, i.e. *transparency of use*, context-aware behaviour is seen as the key enabling factor. Computers already pervade our everyday life - in our phones, fridges, TVs, toasters, alarm clocks, watches, etc - but to fully *disappear*, as in the Weiser’s vision of ubiquitous computing, they have to anticipate the user’s needs in a particular situation and act proactively to provide appropriate assistance. This capability require means to be aware of its surroundings, i.e. context-awareness.

There are many examples of computing systems that are so well implemented that users are not aware that they have interacted with them. Cars are a prime example: ABS (anti-lock braking system) and ESP (Electronic Stability Program) are integrated in cars and influence their usage in extreme situations. Nevertheless, most people will not consciously be thinking of these technologies when operating a car. These technologies are ubiquitous and have disappeared from the user’s conscious mind.

## 14.4 THE NOTION OF CONTEXT

The term context is widely used with very different meanings. The following definitions from the dictionary, as well as the synonyms, provide a basic understanding of the meaning of context.

**context, noun**. *Cause of event*the situation within which something exists or happens, and that can help explain it *Cambridge Dictionary*

**Synonyms for context:** circumstance, situation, phase, position, posture, attitude, place, point, terms, regime, footing, standing, status, occasion, surroundings, environment, location, dependence.

Context-aware computing literature also has several definitions and explanations of what context is. The following are prominent examples that highlight the basic understanding shared in the community.

At the University of Kent, research was conducted that looked at how mobile devices with GPS (externally connected), network access, and further sensors can be used to support the fieldwork of mobile workers. The research team suggested the following definition:

> “ [...] ‘context awareness’, a term that describes the ability of the computer to sense and act upon information about its environment, such as location, time, temperature or user identity. This information can be used not only to tag information as it is collected in the field, but also to enable selective responses such as triggering alarms or retrieving information relevant to the task at hand. ”
>
> *-- Ryan et al 1998*

Figure 14.8: Lancaster castle is one of the locations that was featured in the GUIDE tourist system.



The GUIDE project (GUIDE 2001) at Lancaster University was the first larger and public installation of a research prototype to explore context-awareness in the domain of tourism. It focused on how context can be used to advance a mobile information system for visitors to the historic town of Lancaster. Keith Mitchell suggests the following notion of context in his thesis, based on work with the GUIDE system:

> " [...] two classes of context were identified, namely personal and environmental context. [...]. Examples of environmental context include: the time of day, the opening times of attractions and the current weather forecast. "
>
> -- *Mitchell 2002*

This is an understanding of context where the users themselves are part of the context (e.g. profiles, preferences). Technically, GUIDE followed an interesting approach as it used a modified browser in which context information was used in the background to adapt content and presentation.

Anind Dey has suggested a very generic description of what constitutes context:

> “ Context is any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and application themselves. ”
>
> *-- Dey 2000*


Figure 14.9: This traditional board advertises the same product to all people who pass by – home made soup with bread. In the future such boards will be replaced with digital screens and then it becomes possible to make the content adapt to the current context. If you are interested in the future visions of public display networks have a look at pd-net.org.

For practical purposes, context is often hierarchically structured describing the relevant features. The feature space described by myself (Schmidt et al 1999) is an example of such a structured representation of context, see Figure 14.10. Let us assume you want to design a digital replacement of a menu you find often at the entrance of a restaurant (see Figure 14.9 for an example). If you have a non context-aware version, this would typically show the special of the day. Instead, if you designed it as context-aware, you would want to have different suggestions on the menu depending on who is walking past it. If parents with children walk by, you would show the family-oriented menu; if a couple is looking at it in the evening, you would show the menu for a candle light dinner; and if it is hot and sunny in the afternoon, you would advertise the selection of ice cream you have. A feature space for this application could include the people looking at it, the time of day, and the weather. People could be refined to number of people, age, and gender. The weather could include temperature and whether it rains or not. By providing such a structured space, it becomes easier to link contexts in the real world to adaptations in the system. Try as an example to do a full feature space for the menu and define appropriate adaptations. Even a checklist could be considered as a very simple example of a non-hierarchical feature space.

There is no feature space that is complete and describes all possible options – such a feature space would in fact be an attempt to provide a complete description of the world. The usual approach is to create a feature space for the specific context-aware application that is designed. The advantage of a feature space is that by looking at a set of parameters, it can be easily determined if a situation matches a context or not.

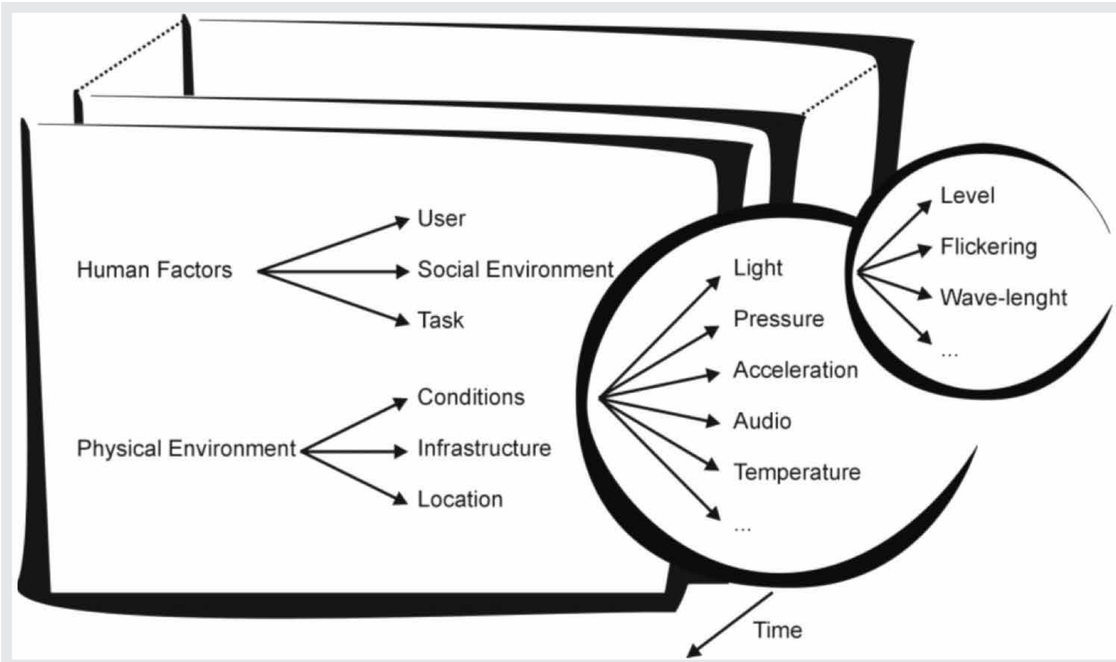



Figure 14.10: Context feature space, detailing light as one feature in the conditions of the physical environment.



Design Hint 1: When building a context-aware system, first create a (hierarchical) feature space with factors that will influence the system behaviour

Knowing which are the factors that should influence the system behaviour, one can start to look at how these factors can be determined in the devices. In many cases this will require sensors that allow the provision of context.

## 14.5 FROM SENSORS TO CONTEXT

The ultimate goal of a context-aware system is for the system to arrive at a representation of the surrounding world that is close to the perception of the user. An important question is how to narrow the gap between the user's and the system's perception (or understanding) of the real world. For location, different means of sensing (e.g. GPS) and interpretation (World Geodetic System, WSG84, post code) are well-established. However, for many other sensors there is typically no single and well-understood way for interpreting the sensed information.

The user's perception of the surroundings is based on human senses, but relates at the same time to experience and memory. Human perception is multifaceted. When walking home from the bus stop late in the evening, a user may perceive that it is dark, quiet, and cold, but at the same time he may perceive the situation as scary. Another user, who was busy the whole day and surrounded by people, may perceive the situation also as dark, quiet, and cold, but at the same time as relaxing and free. This example shows that relying on sensor data alone does not provide the complete picture. It is important to remember that even a perfect design and implementation will not be able to perceive the environment in exactly the same way as the user does.

We now have the following ingredients: The user's perception and the user's experience which both lead to the user's expectations; the system's perceived context drawing from the sensor input; the system's model of the world including a model of the user driving the system's reaction. See Figure 14.5. The main goal of a good and usable design should be to minimize the "awareness mismatch", as illustrated in Figure 14.11.

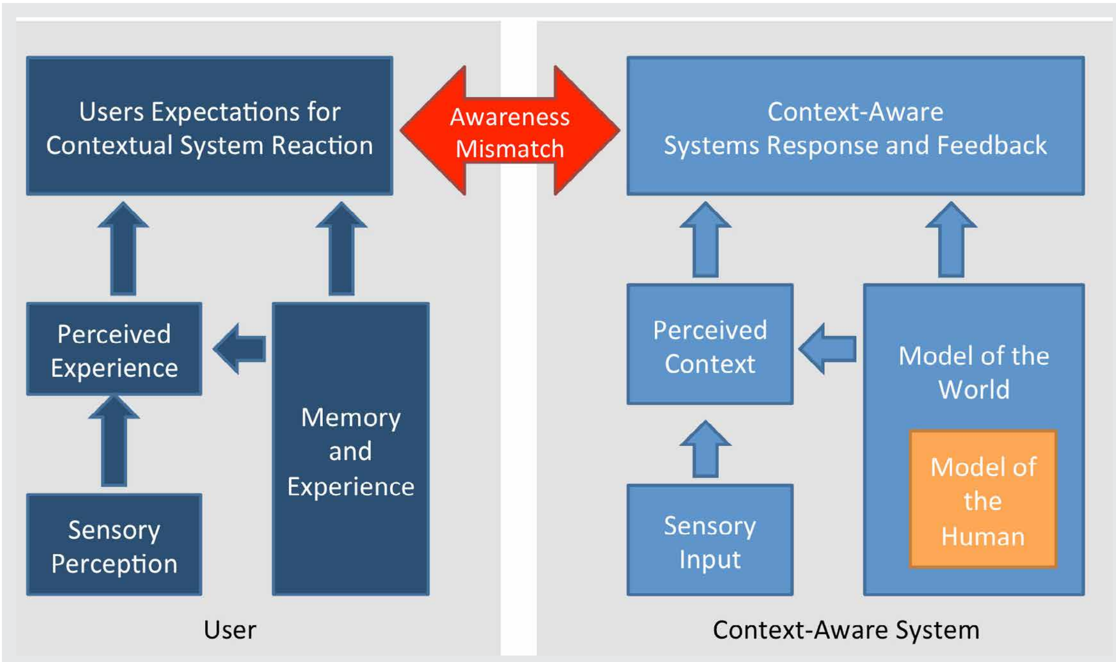



Figure 14.11: The shown User-Context Perception Model (UCPM) highlights the parallel perception processes in the user and in the system. If they are different we create systems with an awareness mismatch, where the system behaviour does not correspond with the users' expectations.



The User-Context Perception Model (UCPM) is a model created to help the designer understand the challenge he faces in creating context-aware systems. It does not describe the way humans work, nor does it prescribe how to implement the system. Nevertheless, the model of a context-aware system, as shown in the model, can provide a good starting point for designing the system architecture of context-aware applications.

By considering the example of a car navigation system, we can look a bit more into the details of the UCPM. If you use a navigation system, you will probably have noticed that it works very well if you are in a city you have never been to before. If you use it around the area where you live, you may, however, sometimes be surprised about the route it tries to direct you to. This phenomenon can be explained with the UCPM. Let's assume the context-aware system (right side in Figure 14.11) is of equal quality in both locations; This means that the difference must be on the user side. The sensory perception (e.g. visual matching of buildings and places you know based on your eyesight) is different in a familiar place and a new place. For the new place, you lack reference points, and the Memory and Experience part in the model differs significantly. In the familiar environment, you will have expectations about which route to take and which way would be a good choice. In the unfamiliar environment, you lack experience and reference points, and hence your expectation is simply that the system will guide you to your destination. The result is that a navigation system that successfully guides you to your destination with a *non-optimal route* will satisfy your expectations in an *unfamiliar* environment, but be frowned upon in a *familiar* environment. A non-optimal route could include taking a detour of a few blocks because the map is out of date, or having to wait at three traffic lights, where you could have alternatively used the slightly longer way over the bridge without traffic lights. In the familiar environment, we have a substantial awareness mismatch, whereas when navigating in new surroundings, we have a minimal awareness mismatch.

Design Hint 2: In the user interface, provide information about the sensory information that is used to determine the context in order to minimize the awareness mismatch.

The quality of context-aware systems, as perceived by the user, is directly related to the awareness mismatch, and a good design aims at designing systems with minimal awareness mismatch. A prerequisite for creating a minimal awareness mismatch is that the user understands what factors have an influence on the system. In the example of a simplistic car navigation system, this factor is only current location and nothing else. In such a case the user knows that the system's reactions are based purely on the current location as well as the destination, and the user may attribute the system's response to these factors. In cases where further parameters play a role - e.g. a navigation system that takes traffic into account - it may be more difficult for the user to understand the causalities behind the system's behaviour. In such an example, the navigation system may suggest different routes in the morning and the evening as the traffic situation is not the same. If the user has no knowledge that the system makes the routing suggestions based on the current location and the traffic situation, it is likely that the user will have a hard time understanding what the system does. As an important rule in the design of context-aware systems, the user should be made aware of the sensory information that the system uses.

There are many examples of devices and applications that provide such feedback, e.g. the type of wireless connectivity in a mobile phone and the symbol for GPS reception. Such hints are essential for the user to understand system behaviour. For example, the user may understand, and accept, that there are significant differences in download speed on a mobile device when supplied with the information that download in some cases happen over the GSM network and in other cases over a WiFi connection. However, if the user has no concept of the difference between a data connection over GSM and WiFi, all this information will not be of much help, and the awareness mismatch remains. Therefore, design hints such as the above-mentioned Design Hint 2 are not absolute rules and do not exempt the designer from doing user studies and usability tests: *Know thy user*, as a popular one-liner goes.

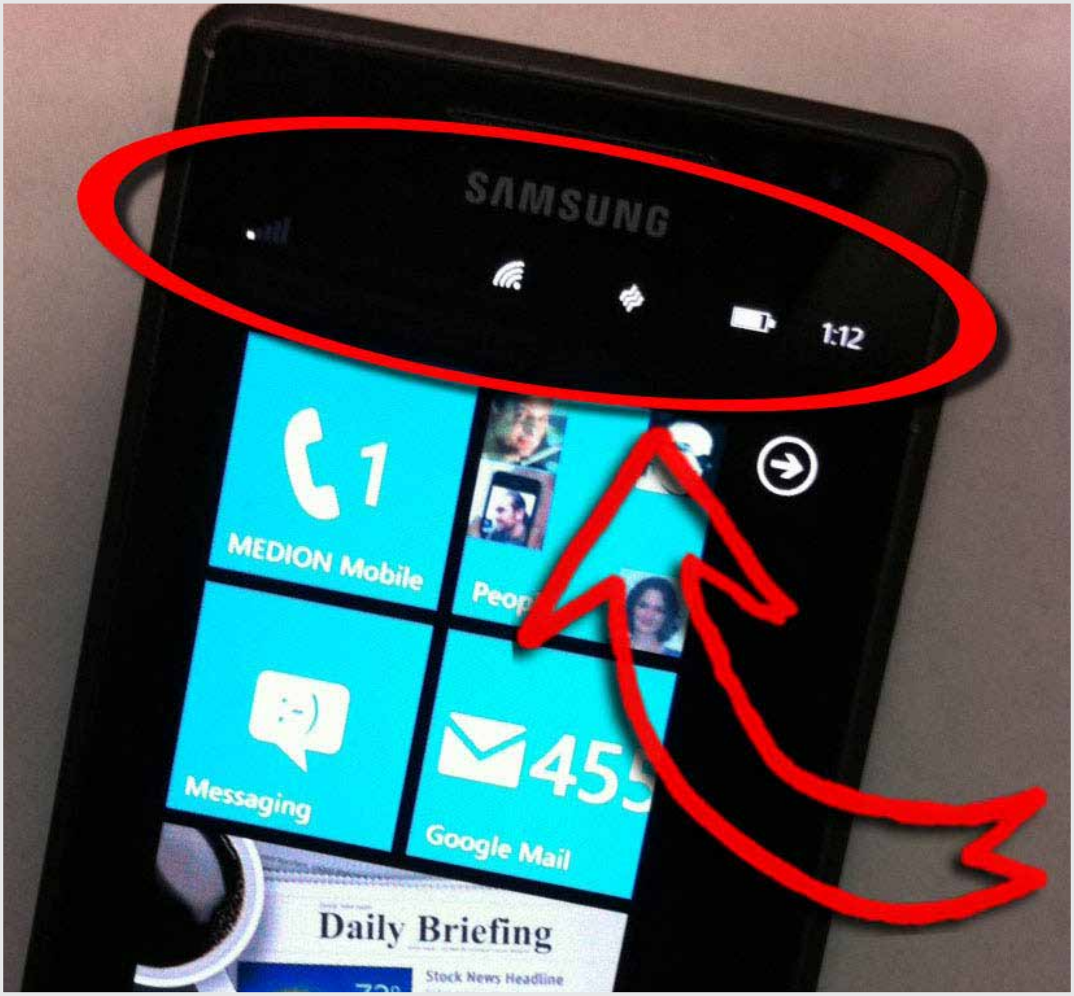



Figure 14.12: Phones provide very simple context-information in the user interface. In this picture the phone has connectivity to the GSM-network as well as to a WiFi base station. Having this information allows the user to better understand system behaviour - for example in the event that the users talks on the phone and the speech quality gets worse after entering a building. Looking at the bars indicating the network strength, the user may realize that the coverage is inadequate. As you have a mental model of the problem and its solution, you move towards a window or back towards the door to regain a satisfactory signal quality.

The very basic idea of sensor-based context awareness is the assumption that similar situations (considered as one context) are represented by similar stimuli. Therefore, sensors may be used to determine contexts based on the assumption that *in similar contexts the sensory input of the characterizing features is similar*. The difficult part is to assess and define what the relevant and characterizing features are. As humans we do this in our everyday activities over and over again. We realize there is meeting in progress when entering a room with people sitting around a table talking - even if we do not know the room and anyone taking part in the meeting. The basic approach is to make an (implicit) analysis of the sensory input received from the surroundings and compare this to situations experienced earlier. Let us assume there is an evening meeting at the University of Stuttgart in the first floor meeting room of the SimTech building, and furthermore that there is a meeting in Lancaster in the InfoLab on the second floor meeting room - both at 10am. Let's compare sensory readings for these two situations and add two further situations: A student lab session in Stuttgart, and cleaning of the meeting room in Lancaster.

| [Situation] | | | | |
|---|---|---|---|---|
| [Feature] | **Meeting Stuttgart** | **Meeting Lancaster** | **Lab Session** | **Cleaning** |
| **Geographic location:** | Stuttgart | Lancaster | Stuttgart | Lancaster |
| **Light on or off:** | light on | light off | light off | light on |
| **Number of people in the room:** | 7 | 9 | 8 | 1 |
| **Language spoken:** | German | English | German | None |
| **Activity in the room:** | sitting | sitting | sitting | movement |
| **Power consumption in the room:** | 2kw | 1,6kw | 3,4kw | 3kw |
| **Devices in use:** | laptop, phone | projector, laptop | laptop, phone | vacuum cleaner |

Table 14.1: Example of situations and their characterizing features.

If we create a matrix in which we count how many features are the same, we arrive at the results in Table 14.2. Using this feature set, we see that the Meeting in Stuttgart is more similar to a Lab Session than to another meeting in Lancaster. If we would choose another set of features, we would get different similarities. This illustrates how important it is to choose the right features for classification. It is important to find the specific features for a context, and in many cases adding further features may be counter-productive.

| [Similarity] | Meeting in S | Meeting in L | Lab Session | Cleaning |
|---|---|---|---|---|
| **Meeting in S** | 7 | 1 | 4 | 1 |
| **Meeting in L** | 1 | 7 | 2 | 1 |
| **Lab Session** | 4 | 2 | 7 | 0 |
| **Cleaning** | 1 | 1 | 0 | 7 |

Table 14.2: Counting similar features for each pair of situations. It becomes clear that just counting any set of features is not going to work well. Choosing the "right" features that are characteristic is essential.

The general approach is to look at which sensor input you expect in a certain context. In Table 14.3, two examples are given. A meeting is detected when several people are present and when these people interact. When the sensor information indicates an ongoing change in light and a certain audio level, as well as an indoor location where the user is stationary, we assume the user is watching TV. The examples show that the expected sensor readings are related to a feature space, described in Figure 14.10. Looking at these descriptions of the expected sensor input, it is apparent that the detection is never perfect. It is easy to create situations that are not a meeting, but classified as a meeting (e.g. having lunch together is likely to be classified as a meeting with the description below). Similarly, we can create a situation that belongs to the context, but is not recognized with the expected sensor input. If the user watches TV while in the garden and perhaps even uses subtitles and has the sound switched off, this would not be recognized.

Such descriptions can be made on very different abstraction levels (e.g. people are present vs. the passive infrared sensor indicating movement). The used descriptions are typically depending on the types of sensors assumed.

| *Context* | *Expected Sensor Input* |
|---|---|
| Meeting | Several people present<br>Interaction between people |
| User watching TV | Light level/color is changing, certain audio level (not silent), type of location is indoors, user is mainly stationary |

Table 14.3: Illustrates example assumptions made for specific sensory inputs on two contexts.

Design Hint 3: Find parameters which are characteristic for a context you want to detect and find means to measure those parameters

In current systems, a wide variety of sensors are used to acquire contextual information. Important sensors used are GPS (for location and speed), light and vision (to detect objects and activities), microphones (for information about noise, activities, and talking), accelerometers and gyroscopes (for movement, device orientation, and vibration), magnetic field sensors (as a compass to determine orientation), proximity and touch sensing (to detect explicit and implicit user interaction), sensors for temperature and humidity (to assess the environment), and air pressure/barometric pressure. There are also sensors to detect the physiological context of the user (e.g. galvanic skin response, EEG, and ECG). Galvanic skin response measures the resistance between two electrodes on the skin. The value measured is dependent on how dry the skin is. Typically, such measurements can be used to determine reactions that change the dryness of the skin, e.g. surprise or fear (lie detectors are based on similar mechanisms). In principle, one can use all types of sensors available on the market to feed the system with context information.

In some applications, it may make sense to use more sensors of the same type to ease the context detection task. For example, to determine the number of speakers and locating their position in a room is straightforward with a set of microphones, whereas this is impossible with a single microphone. The quality of the information gained may also be improved by using a set of sensors rather than one. A simple example is that with a single light sensor one can only detect the light level in the environment, but a larger number of light sensors are the basis for a camera.

To match sensory information with contexts, a matching has to be performed. These perception tasks are typically done by using means of machine learning and data mining. The simplest way is to describe a set of features that define a situation. Then, in any given situation, the system will monitor its sensory input and check if the features match the sensory input. Simple rule-based systems fall into this category. Another example is to record typical situations and calculate representative features for these situations. In a new situation, the features are calculated and compared to the learned (recorded) situations. With a simple so-called “nearest neighbour matching”, the current context can then be calculated.

The quality of the algorithms that calculate the contexts should be assessed to determine how well the system works. These algorithms can be optimized for precision or recall, similarly to classical information retrieval systems. When assessing context-aware systems, it is important to take the probability of a certain context into account; otherwise very rare events may be missed. Assume the following example: You want to build a fire alarm, and you assume that in 10,000 days you will have 1 day where there is a fire. If you pick up a stone from the ground and announce that it is a fire alarm, you can be pretty sure that it does not work as such. Nevertheless, you can still claim that your “fire alarm” will work in 99.99% of the time. However, when providing information for the context “fire (0%)” and the context “non-fire (100%)”, it becomes immediately obvious that a stone is not a useful device for this purpose. To assess this, a confusion matrix is used. It shows the relationship between the real situation and the perceived context for each context defined.

| | | Perceived context by the "system" | |
|---|---|---|---|
| | | Fire | No Fire |
| **Situation in the real world** | **Fire** | 0 % | 100 % |
| | **No Fire** | 0 % | 100 % |

Table 14.4: Confusion matrix for the stone.

| | | Perceived context by the system | |
|---|---|---|---|
| | | Fire | No Fire |
| **Situation in the real world** | **Fire** | 100 % | 0 % |
| | **No Fire** | 0 % | 100 % |

Table 14.5: Confusion matrix for an optimal fire alarm.


FIGURE 14.13: In order to know how well a context detector works, you need to know the recognition performance for each context. In comparison to an actual fire alarm, you most likely agree that a stone will not work well as a fire alarm.

*Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.*

## 14.6 USING CONTEXT IN APPLICATIONS AND USER INTERFACES

When sensory information is available, and a given context can be determined as a result, various functions and behaviours of systems and applications can be linked to contexts. As mentioned earlier, the main motivation for calculating/establishing a given context is to increase a system's understanding of its surrounding environment. This enables the designer to create systems that act differently in different contexts, and if they are well-designed they match the user's expectations in this context, i.e. an *awareness match* cf. Figure 14.11.

Different behaviours can be designed on different levels within a system and range from low-level functionality (e.g. selecting the most appropriate network protocol for the current context), to application behaviour and supported functions (e.g. a mobile device used within the company network may access all company documents, whereas the same device used outside the company may only access a subset of documents), to changes on a user interfaces level (e.g. the zoom level of a map is dependent on the speed with which a car is driving). Often it is hard to clearly discriminate in which category such adaptive behaviour falls.

We generally discriminate between the following types of context-awareness:

- Context-adaptive systems - proactive applications, function triggers, and adaptive applications
- Adaptive and context-aware user interfaces
- Managing interruptions based on situations
- Sharing context and context communication
- Generated data for metadata and implicitly user-generated content
- Context-aware resource management

These basic categories help in the design of context-based applications. In some cases there may not be a clear discrimination between them, or they may be combined in a single application. As early as in the original paper by Schilit et al (1994), a discussion and a table of how context can be used was included. They included a table where they discriminated between what is context-dependent (information or commands) on one side and how context is used (manually or automatically) on the other side. This view of context-aware applications mainly reflects the first and the last types from the above list, i.e. proactive applications and resource management.

It is highly recommended to read this paper by Bill Schilit et al (Schilit et al 1994) as it is the cornerstone and central foundation of context-aware computing. If you are interested in more details on the original work on context-awareness, you may want to read Bill Schilit's PhD thesis.

| | manual | automatic |
|---|---|---|
| information | proximate selection and contextual information | automatic contextual reconfiguration |
| command | contextual commands | context-triggered actions |

Table 14.6: A basic taxonomy of how to create context-aware systems was introduced by Schilit et al in 1994.

### 14.6.1 Context-adaptive systems - proactive applications, function triggers, and adaptive applications

Proactive applications take initiative on behalf of the user, based on the environment and context. An example is a heating system that pro-actively starts heating the house when the context *user on her way home* is detected. A further example is a system that automatically launches a bus timetable application when the user is starting the device at the bus stop. The basic idea of proactive applications is that the system anticipates – based on context – what application the user will need and already executes it. Technically, these are sometimes referred to as triggers. A context *triggers* the launch of an application.

Adaptive applications are conceptually very similar. The main difference is that *functions* are triggered based on context rather than on complete *applications*. However, the granularity of applications and functions can differ greatly; hence the discrimination is not of great relevance. When creating context-adaptive systems, the basic approach is first to define a set of contexts that are relevant for adaptation, then to select a set of functions or applications that are used, and lastly to create a mapping between the contexts and the functionalities. This can be done by making a table like Table 14.7, which applies to an adaptive "home screen." The mapping also defines in what way the trigger is executed. In this example, most of the application functions are executed "On Entry" which means the function is triggered when the user enters the context.

When the user switches to another application, the trigger will not be repeated. The trigger named “Always” will make sure that the triggered application is repeatedly called; in our case, the map will always be shown on the home screen when the user is in the car. In case the user switches to another application, the trigger will check every 30 seconds and switch back to the map (given there is no user activity). A further example is a trigger named “Every 60 Seconds.” Such a trigger is useful to continuously (in this case every 60 seconds) call a function while the user remains in a certain context. The final example of a trigger is “On Exit” which calls a function when the user leaves a context. There is no need for a formal description, but creating a table as shown below helps in the design and implementation of the application. The table also shows that contexts can be present in more than one row as some contexts require more than one function to be triggered. Similarly, functions are not exclusive to one context; a function can be triggered by many contexts.

| Context | Timing/Trigger-Mode | Triggered Function |
|---|---|---|
| **In the office** | On Entry | Show calendar on home screen |
| **On the bus** | On Entry | Run Music Player |
| **In the car** | Always, check every 30 second | Show map on home screen |
| **In the car** | Every 60 Seconds | Submit current location to web service |
| **At home** | On Entry | Show Facebook messages on home screen |
| **At home** | On Exit | Show todo and shopping list |
| **At the gym** | On Entry | Run Music Player |

Table 14.7: Context-Function mapping for a sample application.

Design Hint 4: Designing proactive applications is very difficult because the system has to anticipate what the user wants. In many cases, it is much easier not to present "the application" and rather to present a set of potential interesting applications which the user can launch. For example, a context-aware "home screen" may offer a selection of applications that are useful in a given context rather than attempting to choose the right one.

### 14.6.2 Adaptive and context-aware user interfaces

Context-aware user interfaces are a special case of context-aware functions. Basically, this means a system where the context-aware functions are user interface elements. The level of complexity in adaptation and awareness may differ greatly. A very simple example of a context-aware user interface is the back light of a device that is switched on when the environment is dark. Further examples are audio profiles that suit a particular situation or screen layouts optimized for a given context. On a mobile device, the input modalities may be dependent on the context. For example, in a car a mobile device may use a simple menu with a large font that can be activated by simple voice commands, but the same mobile device may present a more complex user interface when used during a meeting. The paper by Bill Schilit et al (1994), which originally introduced the idea, uses an example of a user interface to select a printer. In Figure 14.14 options are shown how to present the printer selection menu when *proximity* is available as context. This seems an obvious way of presenting the list of printers – but even today, decades later, we do not see this approach in operating systems.

| Name | Room | Distance |
|---|---|---|
| caps | 35-2200 | 200ft |
| claudia | 35-2108 | 30ft |
| perfector | 35-2301 | 20ft |
| snoball | 35-2103 | 100ft |

(a)

| Distance | Name | Room |
|---|---|---|
| 20ft | perfector | 35-2301 |
| 30ft | claudia | 35-2108 |
| 100ft | snoball | 35-2103 |
| 200ft | caps | 35-2200 |

(b)

| Name | Room | Distance |
|---|---|---|
| caps | 35-2200 | 200ft |
| **claudia** | **35-2108** | **30ft** |
| **perfector** | **35-2301** | **20ft** |
| snoball | 35-2103 | 100ft |

(c)

| Name | Room | Distance |
|---|---|---|
| caps | 35-2200 | 200ft |
| claudia | 35-2108 | 30ft |
| perfector | 35-2301 | 20ft |
| snoball | 35-2103 | 100ft |

(d)

FIGURE 14.14: The user interface design show different ways of presenting context (in this case proximity of the printer) to the user; from (Schilit 1994).



Context-aware user interfaces may include both output and input as well as various modalities. Making user interfaces context-aware provides means for creating a user experience that is tailored specifically to each context. This is, however, not as easy as it may sound. Users learn how to use the user interface and adapt their behaviour to it. Users will remember where a menu item is located or which navigation task to perform to get to a specific function. By making the placement of user interface elements adaptive and the structure of user interface dynamic, we make it harder to learn to use a given user interface. If adaptive presentations are not understandable, e.g. by making the underlying causality behind the adaptation clear to the user, it may hinder the user's ability to memorize user interfaces and to interact efficiently.

Taking menu items in a WIMP-style interface as an example, one could argue that menu items should be reordered according to their usage frequency and unused items should disappear. Adaptive menus have been available in earlier Microsoft Office versions, and it proved to be a bad idea – it confused many people and made it much harder to explore and learn to use the product. The subsequent version of Microsoft Office combined stability and context-awareness by keeping all menu items visible but graying out functions that are not available in the current context. This shift worked very well.

Design Hint 5: Use adaptation in the UI with great care and ensure that it is understandable to the user. Good designs maintain stability and support the user in memorizing the UI while using context to reduce complexity

### 14.6.3 Managing interruptions based on situations and sharing context in communication

Interruptions happen all the time. You write an email, and in the middle of the sentence you receive an SMS. Interrupted by the notification tone, you switch your attention to the phone screen and read the message. Then you look back at the email ... and you cannot remember what you wanted to write. Such situations are common, but most of the time they are not critical, and we have learned to cope with them.

Looking at the interruptions, we can state a basic fact: Computers and communication devices are very rude. Imagine you are in line at the library. You are the third in line, and you wait patiently while she deals with the people in front of you. Suddenly a person walks in and goes directly to the front of the queue, cuts off the conversation between the librarian and the person first in line, and asks about the book she has ordered a week ago. This happens rarely, and everyone would be annoyed with the person. When it comes to telecommunication, however, this happens all the time. If two people are having a face-to-face conversation and one of them receives a phone call, that person is very likely to switch her attention to the call and interrupt the face-to-face conversation. This behaviour is perhaps rooted in the old model of synchronous telecommunication where phone calls were expensive and important – which is less true nowadays. Also, before the advent of caller-id, you could not simply return the phone call as you would not know who had called you unless you actually answered the call. Although technology has changed a lot, some of our behaviour around new technologies are still rooted in an understanding of older technology.

There are several ways context can be used to minimize interruptions. For example:

- Context as a source to schedule interruption and communication
- Context sharing to guide the timing of communication

Using context information, we can decide when and how to deliver asynchronous communication. In the example of the SMS interrupting the process of email writing, one can imagine that the notification is postponed until the user presses the send button in the email program. Another option is to change modality, e.g. instead of having the phone deliver the audio notification, we could have a notification like a bubble in the status bar on the computer, which would be less intrusive to the current task. This example shows, however, that there is a trade-off in the design. If we postpone the notification, the user may get the message too late, and if we use an additional or alternative modality, we may still interrupt the user.

In synchronous communication, automation is really hard to achieve. In this case, context sharing is a promising way of using context to improve the user experience. In Schmidt et al (2000), we introduced the idea of publishing one's context to potential calls - e.g. a status like "I am in a meeting" - and leaving the decision to call or not to the caller. The rationale is that only by knowing the caller's context as well as the context of the person to be called, one can decide whether or not the connection should be established. For an introduction of context-aware communication, see (Schmidt et al 2000) and (Schmidt et al 2001).

With the advent of social media services like Facebook and Google+, it has become possible to use context in mainstream communication systems. For example, some phones integrate the address book with Facebook status and location information of the users.

### 14.6.4 Generated Data for Metadata and Implicitly User Generated Content

It is obvious that whatever we do, we do in a context. If we write a paper – we do it somewhere, at some time of the day, after another activity, or before another activity. We may be together with other people while we write it, and we are likely to look up other material while we write. Currently, the text we write does not reflect the context in which it was created. You cannot see that the words you read in this chapter were written during several train journeys. However, if we have context available, we could attach meta-information to each word we write. When later looking at the text, we could look up where it was created and who was present while it was written. One domain for which automatic collection of meta-information is useful is the support of personal memory and personal search. Imagine you look for meeting notes. You may not remember when the meeting took place, but you may remember the place where it was and the people who were present, and that it was late in the evening when it ended. If this meta-information was recorded, you could use it for searching.

Services like YouTube, Wikipedia, Flickr, or Facebook are all examples of media where people generate content and share it. All this content is explicitly generated: videos are recorded, articles are written, and photos are taken – and shared. This explicitly generated content has created a wealth of information and has changed the Web fundamentally.

If we look at context information, there is an equally interesting source of data of user-generated information: *implicitly* user-generated data. If you drive your car from your home to the office, you generate information. Assume you record information from the car (e.g. from the acceleration sensor, vibration sensor, temperature sensor, rain sensor, friction between tires and road) and the navigation system (e.g. speed, location, direction) and share this information on the Web. If a number of people share such context information, it will constitute an entirely new domain of information, ranging from real-time traffic information to road conditions and fine-grained weather reports. Although this scenario is technically feasible, it leaves many questions open with regards to privacy.

### 14.6.5 Context-Aware Resource Management

Managing resources (beyond the user interface) is only indirectly related to the user experience. The basic approach of context-aware resource management is to optimize the operation of a device and its use of resources based on context. One very prominent example is to maximize battery power by using context information, and another one is to switch between available networks based on the current context. They are often realized in lower layers in the operating system.

One should remember that these adaptations may have an impact on the user experience. Transparent resource management is essential for the ease of use of the system; imagine you would have to select the appropriate base station for your mobile phone communication each time you travel to another part of the city. Even if you would just get a message box for each time you register with a new base station, this would render a mobile phone more annoying than useful. This transparency is great as long as it works. If it does not work, or if the user is puzzled about an automatic adaptation done by the system, we should however provide means to inquire into the problem for the user.

This brings us back to a basic design challenge and a trade-off often faced when implementing context-aware systems: finding the balance between visibility and transparency. And this question is often related to visual design as well as abstraction. Take a simple signal strength indicator of a mobile phone: it shows some context information about the resource "connectivity". In this case, the whole information on the quality of the network interface, including package loss, delays, SNR, RSS, etc. are represented by five bars, and this abstraction allows the user to reason, even if he does not know much about wireless radios.

An interesting area in which context information is used as a resource for gaming and for creating interesting experiences is Contextual Gaming, i.e. playing games in different real-world situations. There is a basic introduction to gaming in context and explanations of concrete examples on how to map context and actions in Holleis et al 2011.

## 14.7 IMPLICIT HUMAN COMPUTER INTERACTION

Implicit Human-Computer Interaction, iHCI, (Schmidt 2000) generalizes the concept of context awareness in human-computer interaction. Explicit interaction (traditional, explicit interaction with computers) contradicts the idea of invisible computing and disappearing interfaces. In order to create natural interaction, it appears we need to understand *implicit interaction* in addition to *explicit interaction*. Explicit and implicit interaction can be based on different modalities, including command line, Graphical User Interfaces (GUI), and interaction in the real world. Figure 14.9 outlines an interaction model taking implicit and explicit human computer interaction into account.

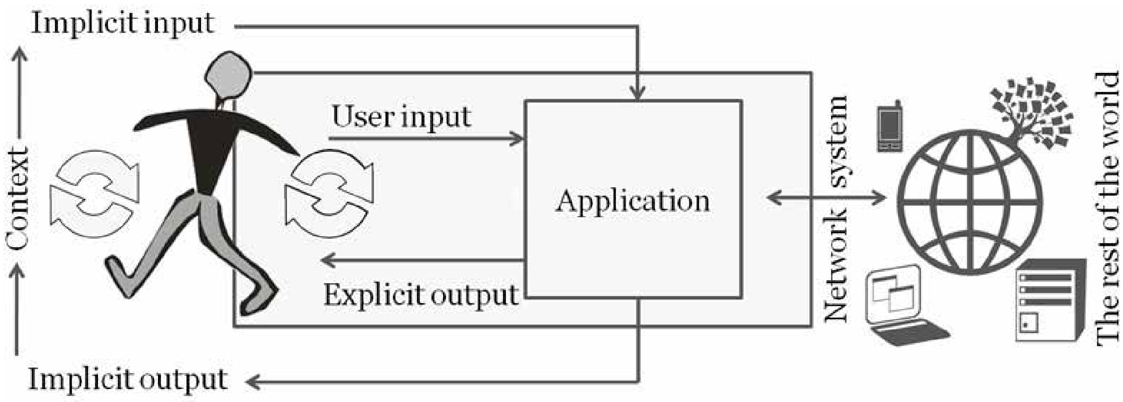



Figure 14.15: This model explains the concept of implicit and explicit human computer interaction, from Schmidt 2000.



Definition: Implicit Human-Computer Interaction (iHCI)

iHCI is the interaction of a human with his environment, including artefacts, with the aim of accomplishing a goal. Within this process, the system acquires implicit input from the user and may present implicit output to the user.

Definition: Implicit Input

Implicit input are actions and behaviour of humans which are done to achieve a goal, and which are not primarily regarded as interaction with a computer, but captured, recognized, and interpreted by a computer system as input.

Definition: Implicit Output

Output of a computer that is not directly related to an explicit input, and which is seamlessly integrated with the environment and the task of the user.

Implicit human-computer interaction is not an alternative to traditional explicit human-computer interaction; it is rather orthogonal. In order to make computers more attentive and natural to use, we need implicit communication channels between humans and computers. Alan Dix used the term *incidental interaction* to describe a similar concept.

## 14.8 SUMMARY AND FUTURE DIRECTIONS

Context-awareness is an exciting and challenging area of human-computer interaction. The basic idea is to give computers perceptual qualities ("eyes and ears") in order to make them recognize the situations in which users interact with information systems. Using sensors, situations can be detected and classified as contexts. Once the system has recognized in which context an interaction takes place, this information can be used to change, trigger, and adapt the behaviour of applications and systems. The input side of implicit human computer interaction looks at information that users generate in order to interact with the real world and thus provides a generalization of context-awareness in human computer interaction.

Creating context-aware interactive systems is hard. One has to keep in mind that users learn how to interact with systems, and that they adapt their behaviour. It is essential that users understand the varying and adaptive behaviour of the application and link it to the situations they are in. Otherwise, they will have a very difficult time learning to use the system. Hence, it is central to create understandable context-aware systems that conform to the users' expectations. In short: Well-designed context-awareness is a great and powerful way to make user-friendly and enjoyable applications. If done wrong, however, context-aware applications may be a source of frustration. Just think of an automatic light, and you will probably come to think of examples that work very well, and others that do not.

Location-awareness as a special form of context have become mainstream as most medium and high-end phones have a GPS receiver and other means for location detection included. Awareness of your friends' and families' whereabouts combined with pedestrian navigation is likely to change the way we coordinate our behaviour, and we will gradually utilize our environment differently as technology changes over the years. Instead of calling someone up to tell that you are late or to ask where to meet, they will have that information readily available (because they are aware of your movements/location). From a design and research perspective, the grand challenge for such a scenario is to create models that allow the individual to control his or her visibility, potentially implicitly, with minimal effort.

With devices that enable us to monitor users in more detail, context-awareness will be included in consumer devices in an ever-increasing degree. Imagine if technologies like cameras and the Kinect, a motion sensing input device by Microsoft for the Xbox 360 video game console, were included into appliances, devices, and your office and home environment. Recognizing where people are and what they do will enable designers to create attentive applications – applications that look at what you do and then react appropriately. The shower will recognize which member of your family is going to use it (e.g. based on the body profile) and pre-select that person's favourite temperature. Designers may explore how appliances can be operated with minimal interaction – potentially just "being there" is enough to work with your environment. Here, a central challenge is to provide means in the user interface to correct wrong choices made by the system, and in a way where the user feels in control.

Another interesting area is implicitly generated content. If we live in sensor-rich environments and with sensor rich devices, we have an unprecedented opportunity to create models of how humans live and interact. Collecting GPS traces to create maps is a start - openstreetmap.org is a good example. If we had similar amounts of information about what people eat, how they sleep, and how much they talk to each other, we would be able to arrive at conclusions like, "people who eat an apple between 5 and 7 in the evening sleep 20% better than people who watch a soap opera." Take some time and think this example through ... I guess you will come up with many new ideas for systems and applications, but at the same time it may also scare you. It is our responsibility as system designers to make a better and more interesting world with the tools and technologies we have at our disposal – and context is an extremely powerful one!

It is exciting to think about how rich sensing and communication will change the way we live. Together with Kristian Kersting and Marc Langheirich, I wrote the article "Perception beyond the here and now" (Schmidt et al 2011). The magazine article discusses how sensor-equipped computing devices are overcoming longstanding temporal and spatial boundaries to human perception.

## 14.9 WHERE TO LEARN MORE

If you want to learn more about context-awareness there are several good resources. A starting point could be chapter 2 and 3 of my PhD dissertation. In these chapters related work is discussed and an approach for context-acquisition is given.

There are several researchers that have worked, and continue to work, on this topic. Their papers may be a good starting point. For early work and foundations (1994-2001) have a look at the publications by Keith Cheverst, Anind Dey, Jason Pascoe, Bill Schilit, and Albrecht Schmidt.

### 14.9.1 Conferences

During the last ten years, the field has become broader, and research on context-awareness that relates to Human Computer Interaction has been published in the following conferences:

**CHI - Human Factors in Computing Systems**

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 2011 | 2010 | 2009 | 2008 | 2007 | 2006 | 2005 | 2004 |
| 2003 | 2002 | 2001 | 2000 | 1999 | 1998 | 1997 | 1996 |

1995 1994 1993 1992 1991 1990 1989 1988
1987 1986 1985 1983 1982

### UbiComp - International Conference on Ubiquitous Computing

2012 2011 2010 2008 2007 2006 2005 2004
2003 2002 2001 2000 1999

### PerCom - IEEE International Conference on Pervasive Computing and Communications

2008 2007 2007 2006 2006 2005 2005 2004
2004 2003

### Pervasive - International Conference on Pervasive Computing

2008 2008 2007 2007 2006 2005 2005 2004
2002

### LoCA - Symposium on Location and Context Awareness

2009 2007 2006 2005

## 14.9.2 Journals

Springer Personal and Ubiquitous Computing

2009 2008 2007 2006 2005 2004 2003 2002
2001 2000 1999 1998 1997

### IEEE Pervasive Computing

2012 2010 2008 2009 2008 2007 2006 2005
2004

### 14.9.3 Other resources

If you are interested in thinking about, or designing, the future of sensing and context there are many directions to look. One area is reality mining, which relates very closely to implicit content generation as discussed above: For an introduction see reality.media.mit.edu/, and publications can be found at reality.media.mit.edu/publications

## 14.10 COMMENTARY BY KEITH CHEVERST

How to cite this commentary in your report

### Keith Cheverst

© *Keith Cheverst*

Keith Cheverst is a Senior Lecturer in the Computing Department at Lancaster University and has recently held positions as Visiting Research Scientist at Melbourne University and Microsoft Research, Cambridge. For the last 10 years he has participated actively in the areas of Mobile computing, CSCW, HCI and Ubiquitous computing (with his interest in the latter area focussing on the study...

Keith Cheverst
Keith Cheverst is a member of The Interaction Design Foundation

Context-aware behaviour is standard on modern smart phones and for me it's a little strange (and encouraging) to think that a short ten or so years ago we were still writing research papers on the subject. Albrecht has been an eminent figure on context-awareness during this time and this chapter effortlessly captures both the depth and breadth of the subject matter in a text that is refreshingly easy to digest.

I sneakily used a draft of the chapter as the reading material for a master's level HCI class recently and it worked very well indeed. The chapter provides an excellent historical perspective on the subject matter including the key motivating forces behind context-awareness, namely mobile and ubiquitous computing. It was good to be reminded of the early examples introduced by Bill Schilit such as the listing of printers according to their proximity - something I would still like to see. But Albrecht's own examples presented in the chapter helped to both illustrate key concepts and stimulate much interesting design discussion.

One important key concept that Albrecht addresses very clearly and which students can struggle with is the hierarchical model by which raw sensor data can be translated into higher level context triggers. While working through the detailed worked example in this section, I found myself musing on how familiar my current class of students are with sensors, such as accelerometers and gyroscopes, compared to a class of just a few years ago.

In addition to its technical/architectural treatment of context-awareness the chapter is also rich in its coverage of 'implications for HCI'. As Albrecht argues clearly, context-awareness provides a powerful tool for the interaction designer. When used well it enables the design of systems that can help 'take the load off' but when used poorly can lead to the production of systems that prove burdensome to say the least.

One of the key challenges faced by any designer wishing to implement context-aware behaviour is that of maintaining predictability and I found myself nodding in agreement while reading Albrecht's discussion of what he terms 'the awareness mismatch' which he succinctly models using the User-Context Perception Model (UCPM).

Another challenging issue for the designer is how best to maintain appropriate levels of control for the user, to keep the user in the loop so to speak. This is especially important for those context-aware systems that implement proactive/adaptive behaviour. As Albrecht states "...a central challenge is to provide means in the user interface to correct wrong choices made by the system, and in a way where the user feels in control". I am reminded very much of an annoying 'habit' exhibited by my current 'smart' phone. I get frustrated when it insists on pausing music playback whenever I place the thing face down. I often place it face down because the speaker is on the back of the phone and so by placing it face down (i.e. 'speaker up') I achieve the most volume. At times like this I wish there was some kind of 'hold' button that would cause the phone to enter a mode whereby all context triggers based on physical actions would be disabled (or rather banished). However, for whatever reason (patent?) the phone doesn't have one and so I don't feel as if I have this control (maybe there is such a feature described somewhere in the manual...). Next to the hold button I would have a 'show me the rule explaining why you just did that' button. As Albrecht states: "...if the user is puzzled about an automatic adaptation done by the system, we should however provide means to inquire into the problem for the user". This reminds me of some of the research by Judy Kay on scrutability (Kay, 1998) within the user modelling domain. At Lancaster we explored this issue with the development of a system that enabled the user to scrutinise (and possibly override) rules inferred by a proactive context-aware system (supporting office environment control) based on context history (Cheverst, 2005). Supporting such user inquiry while maintaining simplicity is clearly hard (and not a simple matter of adding more buttons...). But unchecked the feeling of 'why is IT doing that and how can I stop it!?" can start to feel like a war of attrition with a contrary being. Indeed, when summarising context-aware behaviour, Albrecht states "The basic idea is to give computers perceptual qualities ("eyes and ears") in order to make them recognize the situations in which users interact with information systems". So my phone needs to be smart enough to recognise and discern between the situation where I place it face down because I want greater volume and the situation where I quickly place the phone face down because the phone has started to ring during a meeting and I want it to go silent immediately. Albrecht provides a wonderfully clear worked example to demonstrate how one can procedurally think about the appropriate matching of sensory-based inputs to certain situations, i.e. "to assess and define what the relevant and characterizing features are". So hopefully future designers of context-aware behaviour will take the opportunity to read Albrecht's chapter and produce the kind of smart behaviours that assist rather than hinder and befriend rather than estrange.

### References


- Kay, J.: A scrutable user modelling shell for user-adapted interaction, PhD Thesis, Basser Department of Computer Science, University of Sydney, Australia (1998)
- Cheverst, K, Byun, H, Fitton, D, Sas, C, Kray, C & Villar, N, 'Exploring Issues of User Model Transparency and Proactive Behaviour in an Office Environment Control System', *Special Issue of UMUAI (User Modelling and User-Adapted Interaction) on User Modeling in Ubiquitous Computing,* pp. 235-273. (2005)


## 14.11 REFERENCES


*Dey, Anind (2000). Providing Architectural Support for Building Context-Aware Applications - Ph. D. Thesis Dissertation. College of Computing, Georgia Tech http://www.cc.gatech.edu/fce/ctk/pubs/dey-thesis.pdf*

*GUIDE (2011). The GUIDE Project Home Page - at Lancaster University. Retrieved 19 January 2011 from GUIDE: http://www.guide.lancs.ac.uk/overview.html*

*Mitchell, Keith (2002). Supporting the Development of Mobile Context-Aware Computing - Ph.D. Thesis. Department of Computing, Lancaster University*

*Ryan, Nick S., Pascoe, Jason and Morse, David R. (1998): Enhanced Reality Fieldwork: the Context-aware Archaeological Assistant. In: Gaffney, V., Leusen, M. van and Exxon, S.*

*(eds.). "Computer Applications in Archaeology - British Archaeological Reports". Oxford: Tempus Reparatum*

*Schilit, Bill N., Adams, Norman I. and Want, Roy (1994): Context-Aware Computing Applications. In: Proceedings of the Workshop on Mobile Computing Systems and Applications December, 1994, Santa Cruz, CA, USA.*

*Schmidt, Albrecht (2000): Implicit Human Computer Interaction Through Context. In Personal and Ubiquitous Computing, 4 (2)*

*Schmidt, Albrecht, Beigl, Michael and Gellersen, Hans-Werner (1999): There is more to context than location. In Computers & Graphics, 23 (6) pp. 893-901*

*Schmidt, Albrecht, Takaluoma, Antti and Mäntyjärvi, Jani (2000): Context-Aware Telephony Over WAP. In Personal and Ubiquitous Computing, 4 (4) pp. 225-229*

*Schmidt, Albrecht, Langheinrich, Marc and Kersting, Kristian (2011): Perception beyond the Here and Now. In IEEE Computer, 44 (2) p. 86–88*

*Schmidt, Albrecht, Stuhr, Tanjev and Gellersen, Hans-Werner (2001): Context-Phonebook - Extending Mobile Phone Applications with Context. In: 3rd Mobile Human-Computer Interaction Workshop 2001.*

*Weiser, Mark (1991): The Computer for the 21st Century. In Scientific American, 265 (3) pp. 94-104*

## About the author

### Albrecht Schmidt

© *Albrecht Schmidt*

Personal Homepage:
albrecht-schmidt.blogspot.com/

Current place of employment:
University of Stuttgart

I am a professor at the University of Stuttgart. My central research interests are novel user interfaces and innovative applications enabled by ubiquitous computing. Before moving to Stuttgart I was a professor at the Univeristy of Duisburg-Essen, had a joined position between the Fraunhofer Institute for Intelligent Analysis and Information Systems (IAIS) and the University of Bonn. I studied computer science in Ulm and Manchester and afterwards worked as a researcher at the University of Karlsruhe and at Lancaster University. There I completed in 2003 my PhD thesis on the topic of “Ubiquitous Computing - Computing in Context”. Before I became professor at the B-IT-Center I headed the DFG-funded “Embedded Interaction Research Group” at the Ludwig-Maximilians University in Munich. My teaching and research interests are in media informatics and in particular in the areas of user interface engineering, pervasive computing and mobile interactive systems.

## YOUR NOTES AND THOUGHTS ON CHAPTER 14

Record your notes and thoughts on this chapter. If you want to share these thoughts with others online, go to the bottom of the page at:
http://www.interaction-design.org/encyclopedia/context-aware_computing.html

**NOTES:**

# Chapter 15

# Usability Evaluation

***by Gilbert Cockton.***

Put simply, usability evaluation assesses the extent to which an interactive system is easy and pleasant to use. Things aren't this simple at all though, but let's start by considering the following propositions about usability evaluation:

1. Usability is an inherent measurable property of all interactive digital technologies
2. Human-Computer Interaction researchers and Interaction Design professionals have developed evaluation methods that determine whether or not an interactive system or device is usable.
3. Where a system or device is usable, usability evaluation methods also

determine the extent of its usability, through the use of robust, objective and reliable metrics

4. Evaluation methods and metrics are thoroughly documented in the Human-Computer Interaction research and practitioner literature. People wishing to develop expertise in usability measurement and evaluation can read about these methods, learn how to apply them, and become proficient in determining whether or not an interactive system or device is usable, and if so, to what extent.

The above propositions represent an ideal. We need to understand where current research and practice fall short of this ideal, and to what extent. Where there are still gaps between ideals and realities, we need to understand how methods and metrics can be improved to close this gap. As with any intellectual endeavour, we should proceed with an open mind, and acknowledge that not only are some or all of the above propositions not true, but that they can never be so. We may have to close some doors here, but in doing so, we will be better equipped to open new ones, and even go through them.

## 15.1 FROM FIRST WORLD OPPRESSION TO THIRD WORLD EMPOWERMENT

Usability has been a fundamental concept for Interaction Design research and practice, since the dawn of Human-Computer Interaction (HCI) as an inter-disciplinary endeavour. For some, it was and remains HCI's core concept. For others, it remains important, but only as one of several key concerns for interaction design.

It would be good to start with a definition of usability, but we are in contended territory here. Definitions will be presented in relation to specific positions on usability. You must choose a one that fits your design philosophy. Three alternative definitions are offered below.

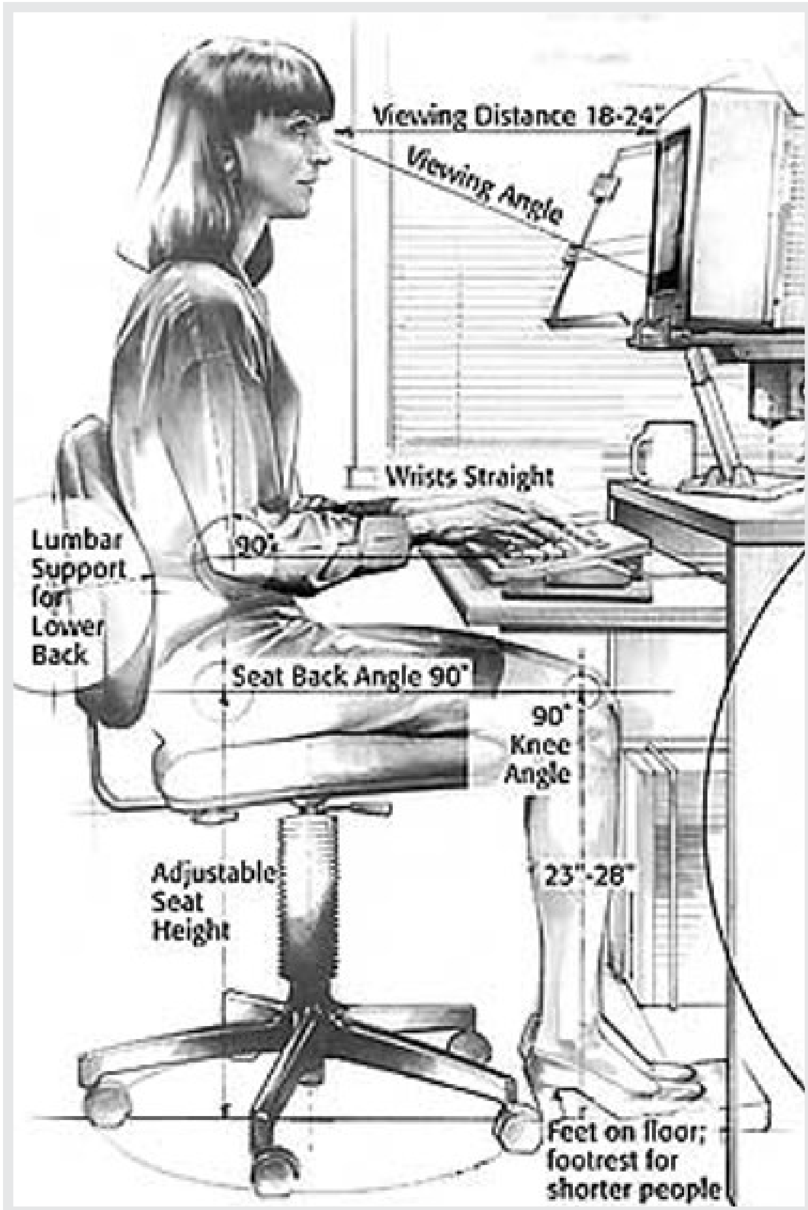
It would also be good to describe how usability is evaluated, but alternative understandings of usability result in different practices. Professional practice is very varied, and much does not generalise from one project to the next. Evaluators must choose how to evaluate. Evaluations have to be designed, and designing requires making choices.

### 15.1.1 The Origins of HCI and Usability

HCI and usability have their origins in the falling prices of computers in the 1980s, when for the first time, it was feasible for many employees to have their own *personal computer* (a.k.a PC). For their first three decades of computing, almost all users were highly trained specialists of expensive centralised equipment. A trend towards less well trained users began in the 1960s with the introduction of timesharing and minicomputers. With the use of PCs in the 1980s, computer users increasingly had no, or only basic, training on operating systems and applications software. However, software design practices continued to implicitly assume knowledgeable and competent users, who would be familiar with technical vocabularies and systems architectures, and also possess an aptitude for solving problems arising from computer usage. Such implicit assumptions rapidly became unacceptable. For the typical user, interactive computing became associated with constant frustrations and consequent anxieties. Computers were obviously too hard to use for most users, and often absolutely unusable. *Usability* thus became a key goal for the design of any interactive software that would not be used by trained technical computer specialists. Popular terms such as “user-friendly” entered everyday use. Both usability and user-friendliness were initially understood to be a property of interactive software. Software either was usable or not. Unusable software could be made usable through re-design.


IBM
IBM
IBM


*Courtesy of Jeremy Banks. Copyright: CC-Att-2 (Creative Commons Attribution 2.0 Unported).*




*Courtesy of Berkeley Lab. Copyright: pd (Public Domain (information that is common property and contains no original authorship)).*

Figure 15.1 A-B-C: The Home Personal Computer (PC) and Associated Peripherals is Now an Everyday Sight in Homes Worldwide. Usability became a critical issue with PC's introduction.

### 15.1.2 From Usability to User Experience via Quality in Use

During the 1990s, more sophisticated understandings of usability shifted from an all-or-nothing binary property to a continuum spanning different extents of usability. At the same time, the focus of HCI shifted to *contexts of use* (Cockton 2004). Usability ceased to be HCI's dominant concept, with research increasingly focused on the fit between interactive software and its surrounding usage contexts. Quality in use no longer appeared to be a simple issue of how inherently usable an interactive system was, but how well it fitted its context of use. *Quality in use* became a preferred alternative term to *usability* in international standards work, since it avoided implications of usability being an absolute context-free invariant property of an interactive system. Around the turn of the century, the rise of networked digital media (e.g., web, mobile, interactive TV, public installations) added novel emotional concerns for HCI, giving rise to yet another more attractive term than usability: *user experience*.

Current understandings ofusability are thus different from those from the early days of HCI in the 1980s. Since then, ease of use has improved though both attention to interaction design and improved levels of IT literacy across much of the population in advanced economies. Familiarity with basic computer operations is now widespread, as evidenced by terms such as "digital natives" and "digital exclusion", which would have had little traction in the 1980s. Usability is no longer automatically the dominant concern in interaction design. It remains important, with frustrating experiences of difficult to use digital technologies still commonplace. Poor usability is still with us, but we have moved on from Thomas Landauer's 1996 *Trouble with Computers* (Landauer 1996). When PCs, mobile phones and the internet are instrumental in major international upheavals such as the Arab Spring of 2011, the value of digital technologies can massively eclipse their shortcomings.

### 15.1.3 From Trouble with Computers to Trouble from Digital Technologies

Readers from developing countries can today experience Landauer's *Trouble with Computers* as the moans of oversensitive poorly motivated western users. On 26th January 1999, a "hole in the wall" was carved at the NIIT premises in New Delhi. Through the this hole, a freely accessible computer was made available for people in the adjoining slum of Kalkaji. It became an instant hit, especially with children who, with no prior experience, learnt to use the computer on their own. This prompted NIIT's Dr. Mitra to propose the following hypothesis:

> " The acquisition of basic computing skills by any set of children can be achieved through incidental learning provided the learners are given access to a suitable computing facility, with entertaining and motivating content and some minimal (human) guidance "
>
> -- http://www.hole-in-the-wall.com/Beginnings.html

There is a strong contrast here with the usability crisis of the 1980s. Computers in 1999 were easier to use than those from the 1980s, but they still presented usage challenges. Nevertheless, residual usability irritations have limited relevance for this century's slum children in Kalkaji.

The world is complex, what matters to people is complex, digital technologies are diverse. In the midst of this diverse complexity, there can no simple day of judgement when digital technologies are sent to usability heaven or unusable hell.

The story of usability is a perverse journey from simplicity to complexity. Digital technologies have evolved so rapidly that intellectual understandings of usability have never kept pace with the realities of computer usage. The pain of old and new world corporations struggling to secure returns on investment in IT in the 1980s has no rendezvous with the use of social media in the struggles for democracy in third world dictatorships. Yet we cannot simply discard the concept of usability and move on. Usage can still be frustrating, annoying, unnecessarily difficult and even impossible, even for the most skilled and experienced of users.


Figure 15.2: NIIT's "hole in the wall" Computer in New Delhi.

Figure 15.3: NIIT's "hole in the wall" Computer in New Delhi.



### 15.1.4 From HCI's sole concern to an enduring important factor in user experience

This encyclopaedia entry is not a requiem for usability. Although now buried under broader layers of *quality in use* and *user experience*, usability is not dead. For example, I provide some occasional IT support to my daughter via SMS. Once, I had to explain how to force the restart of a recalcitrant stalled laptop. Her last message to me on her problem was:

> "It's fixed now! I didn't know holding down the power button did something different to just pressing it."

Given the hidden nature of this functionality (a short press hibernates many laptops), it is no wonder that my daughter was unaware of the existence of a longer ‘holding down’ action. Also, given the rare occurrences of a frozen laptop, my daughter would have had few chances to learn. She had to rely on my knowledge here. There is little she could have known herself without prior experience (e.g., of iPhone power down).

Figure 15.4: Holding or Pressing? Who’s to Know?.

*Courtesy of Rico Shen. Copyright: CC-Att-SA-3 (Creative Commons Attribution-ShareAlike 3.0).*

The enduring realities of computer use that usability seeks to encompass remain real and no less potentially damaging to the success of designs today than over thirty years ago. As with all disciplinary histories, the new has not erased the old, but instead, like geological strata, the new overlies the old, with outcrops of usability still exposed within the wider evolving landscape of user experience. As in geology, we need to understand the present intellectual landscape in terms of its underlying historical processes and upheavals.

What follows is thus not a journey through a landscape, but a series of excavations that reveal what usability has been at different points in different places over the last three decades. With this in place, attention is refocused on current changes in the interaction design landscape that should give usability a stable place within a broader understanding of designing for human values (Harper et al. 2008). But for now, let us begin at the beginning, and from there take a whistle stop tour of HCI history to reveal unresolved tensions over the nature of usability and its relation to interaction design.

## 15.2 FROM USABILITY TO USER EXPERIENCE - TENSIONS AND METHODS

The need to design interactive software that could be used with a basic understanding of computer hardware and operating systems was first recognised in the 1970s, with pioneering work within software design by Fred Hansen from Carnegie Mellon University (CMU), Tony Wasserman from University of California, San Francisco (UCSF), Alan Kay from Xerox Palo Alto Research Center (PARC), Engel and Granda from IBM, and Pew and Rollins from BBN Technologies (for a review of early HCI work, see Pew 2002). This work took several approaches, from detailed design guidelines to high level principles for both software designs and their development processes. It brought together knowledge and capabilities from psychology and computer science. The pioneering group of individuals here was known as the *Software Psychology Society*, beginning in 1976 and based in the Washington DC area (Shneiderman 1986). This collaboration between academics and practitioners from cognitive psychology and computer science forged approaches to research and practice that remained the dominant paradigm in Interaction Design research for almost 20 years, and retained a strong hold for a further decade. However, this collaboration contained a tension on the nature of usability.

The initial focus was largely cognitive, focusing on causal relationships between user interface features and human performance, but with different views on how user interface features and human attributes would interact. If human cognitive attributes are fixed and universal, then user interface features can be inherently usable or unusable, making usability an *inherent binary property* of interactive software, i.e., an interactive system simply *is* or *is not* usable. Software could be inherently usable by conformance to guidelines and principles that could be discovered, formulated and validated by psychological experiments. However, if human cognitive attributes vary not only between individuals, but across different settings, then usability becomes an emergent property that depends, not only on features and qualities of an interactive system, but also on who was using it, and on what they were trying to do with it. The latter position was greatly strengthened in the 1990s by the "turn to the social" (Rogers et al. 1994). However, much of the intellectual tension here was defused as HCI research spread out across a range of specialist communities focused on the Association for Computing Machinery's conferences such as the ACM Conference on Computer Supported Cooperative Work (CSCW) from 1986 or the ACM Symposium on User Interface Software and Technology (UIST) from 1988. Social understandings of usability became associated with CSCW, and technological ones with UIST.

Psychologically-based research on usability methods in major conferences remained strong into the early 1990s. However, usability practitioners became dissatisfied with academic research venues, and the first UPA (Usability Professionals Association) conference was organised in 1992. This practitioner schism happened only 10 years after the Software Psychology Society had co-ordinated a conference in Gaithersburg, from which the ACM CHI conference series emerged. This steadily removed much applied usability research from the view of mainstream HCI researchers. This separation has been overcome to some extent by the UPA's open access *Journal of Usability Studies*, which was inaugurated in 2005.

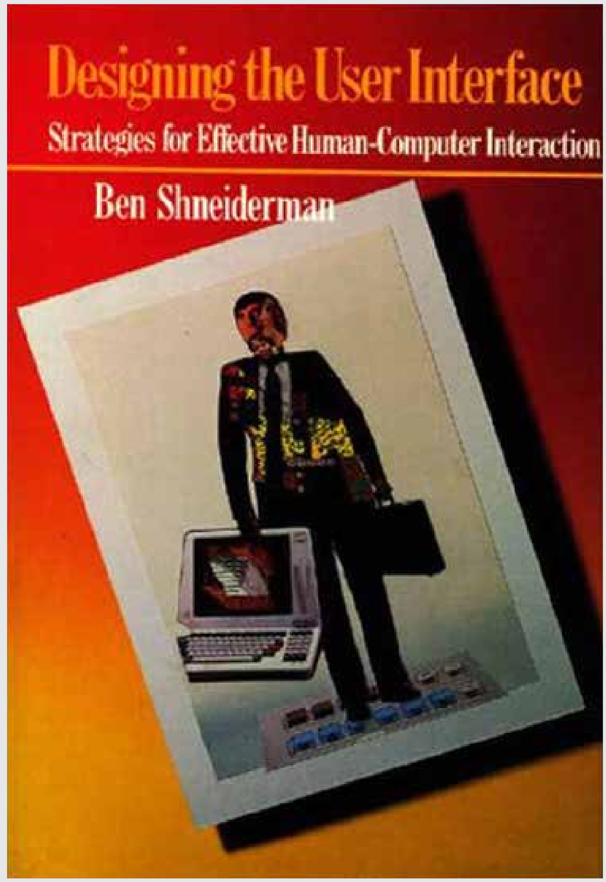



FIGURE 15.5: Ben Shneiderman, Software Psychology Pioneer, Authored the First HCI Textbook.



### 15.2.1 New Methods, Damaged Merchandise and a Chilling Fact

There is thus a dilemma at the heart of the concept of usability: is it a property of systems or a property of usage? A consequence of 1990s fragmentation within HCI research was such important conceptual issues were brushed aside in favour of pragmatism amongst those researchers and practitioners who retained a specialist interest in usability. By the early 1990s, a range of methods had been developed for evaluating usability. *User Testing* (Dumas and Redish 1993) was well established by the late 1980s, as essentially a variant of psychology experiments with only dependent variables (the interactive system being tested became the independent *constant*). *Discount methods* included rapid low cost user testing, as well as inspection methods such as *Heuristic Evaluation* (Nielsen 1994). Research on *model-based methods* such as the *GOMS* model (Goals, Operators, Methods, and Selection rules - John and Kieras 1996) continued, but with mainstream publications becoming rarer by 2000.

With a choice of inspection, model-based and empirical (e.g., user testing) evaluation methods, questions arose as to which evaluation method was best and when and why. Experimental studies attempted to answer these questions by treating evaluation methods as independent variables in comparison studies that typically used problem counts and/or problem classifications as dependent variables. However, usability methods are too incompletely specified to be consistently applied, letting Wayne Gray and Marilyn Salzman invalidate several key studies in their *Damaged Merchandise* paper of 1998. Commentaries on their paper failed to undo the damage of the Damaged Merchandise charge, with further papers in the first decade of this century adding more concerns over not only method comparison, but the validity of usability methods themselves. Thus in 2001, Morten Hertzum and Niels Jacobsen published their "chilling fact" about use of usability methods: there are substantial evaluator effects. This should not have surprised anyone with a strong grasp of Gray and Salzman's critique, since inconsistencies in usability method use make valid comparisons close to impossible in formal studies, and they are even more extensive in studies that attempt no control.

Critical analyses by Gray and Salzman, and by Hertzum and Jacobsen, made pragmatic research on usability even less attractive for leading HCI journals and conferences. The method focus of usability research shrunk, with critiques exposing not only the consequences of ambivalence over the causes of poor usability (system, user or both?), but also the lack of agreement over what was covered by the term usability.

Figure 15.6: 2020 Usability Evaluation Method Medal Winners.

*Courtesy of kinnigurl. Copyright: CC-Att-SA-2 (Creative Commons Attribution-ShareAlike 2.0 Unported).*

### 15.2.2 We Can Work it Out: Putting Evaluation Methods in their (Work) Place

Research on *usability* and *methods* has since the late 00s been superseded by research on *user experience* and *usability work*. *User experience* is a broader concept than usability, and moves beyond efficiency, task quality and vague user satisfaction to a wide consideration of cognitive, affective, social and physical aspects of interaction.

*Usability work* is the work carried out by usability specialists. Methods contribute to this work. Methods are not used in isolation, and should not be assessed in isolation. Assessing methods in isolation ignores the fact thatusability work combines, configures and adapts multiple methods in specific project or organisational contexts. Recognition of this fact is reflected in an expansion of research focus from *usability methods* to *usability work,* e.g., is in PhDs (Dominic Furniss, Tobias Uldall-Espersen, Mie Nørgaard) associated with the European MAUSE project (COST Action 294, 2004-2009). It is also demonstrated in the collaborative research of MAUSE Working Group 2 (Cockton and Woolrych 2009).

A focus on actual work allows realism about design and evaluation methods. Methods are only one aspect of usability work. They are not a separate component of usability work that has deterministic effects, i.e., effects that are guaranteed to occur and be identical across all project and organisational contexts. Instead, broad evaluator effects are to be expected, due to the varying extent and quality of design and evaluation resources in different development settings. This means that we cannot and should not assess usability evaluation methods in artificial isolated research settings. Instead, research should start with the concrete realities of usability work, and within that, research should explore the true nature of evaluation methods and their impact.

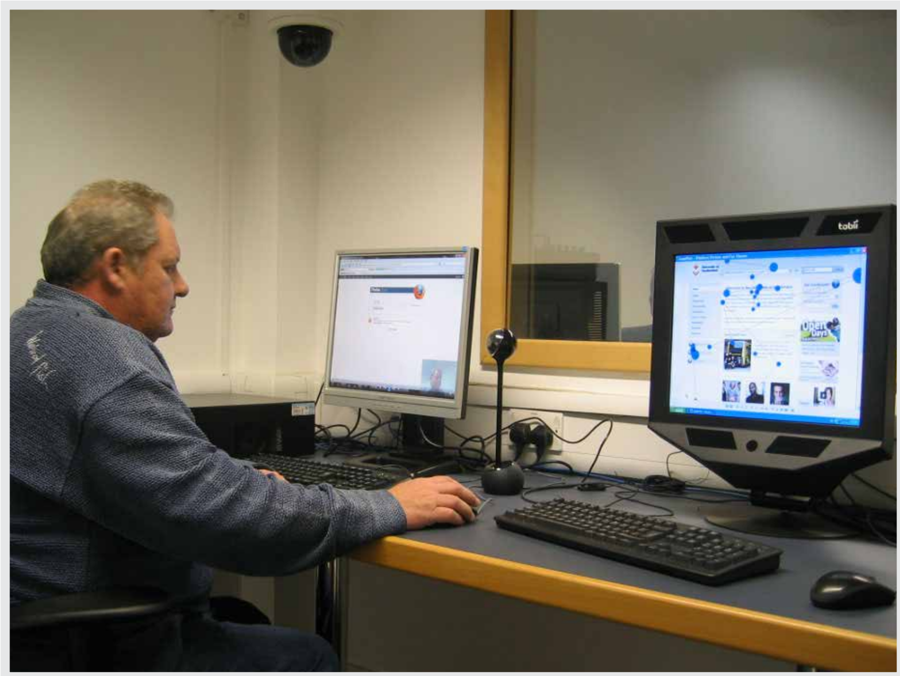



Figure 15.7: Usability Expert at Work: Alan Woolrych at the University of Sunderland using a minimal Mobile Usability Lab setup of webcam with audio recording plus recording of PC screen and sound, complemented by an eye tracker to his right.



### 15.2.3 The Long and Winding Road: Usability's Journey from Then to Now

Usability is now one aspect of user experience, and usability methods are now one loosely pre-configured area of user experience work. Even so, usability remains important. The value of the recent widening focus to user experience is that it places usability work in context. Usability work is no longer expected to establish its value in isolation, but is instead one of several complementary contributors to design quality.

Usability as a core focus within HCI has thus passed through phases of psychological theory, methodological pragmatism and intellectual disillusionment. More recent foci on quality in use and user experience make it clear that Interaction Design cannot just focus on features and attributes of interactive software. Instead, we must focus on the interaction of users and software in specific settings. We cannot reason solely in terms of whether software is inherently usable or not, but instead have to consider what does or will happen when software is used, whether successfully, unsuccessfully, or some mix of both. Once we focus on interaction, a wider view is inevitable, favouring a broad range of concerns over a narrow focus on software and hardware features.

*Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.*



FIGURE 15.8 A-B: What's Sailable: A Boat Alone or a Crewed Boat in Specific Sea Conditions? A Similar Question Arises for Usable Systems.

Many of the original concerns of 1980s usability work are as valid today as they were 30 years ago. What has changed is that we no longer expect usability to be the only, or even the dominant, human factor in the success of interactive systems. What has not changed is the potential confusion over what usability is, which has existed from the first days of HCI, i.e., whether software or usage is usable. While this may feel like some irrelevant philosophical hair-splitting, it has major consequences for usability evaluation. If software can be inherently usable, then usability can be evaluated solely through direct inspection. If usability can only be established by considering usage, then indirect inspection methods (walkthroughs) or empirical user testing methods must be used to evaluate.

### 15.2.4 Usability Futures: From Understanding Tensions to Resolving Them

The form of the word ‘usability’ implies a property that requires an *essentialist* position, i.e., one that sees properties and attributes as been inherent in objects, both natural and artificial (in Philosophy, this is called an essentialist or substantivist *ontology*). A literal understanding of usability requires interactive software to be inherently usable or unusable. Although a more realistic understanding sees usability as a property of interactive use and not of software alone, it makes no sense to talk of use as being usable, just as it makes no sense to talk of eating being edible. This is why the term *quality in use* is preferred for some international standards, because this opens up a space of possible qualities of interactive performance, both in terms of what is experienced, and in terms of what is achieved, for example, an interaction can be ‘successful’, ‘worthwhile’, ‘frustrating’, ‘unpleasant’, ‘challenging’ or ‘ineffective’.

Much of the story of usability reflects a tension between the tight software view and the broader sociotechnical view of system boundaries. More abstractly, this is a tension between substance (essence) and relation, i.e., between *inherent qualities* of interactive software and *emergent qualities* of interaction. In philosophy, the position that relations are more fundamental than things in themselves characterises a *relational ontology*.

Ontologies are theories of being, existence and reality. They lead to very different understandings of the world. Technical specialists and many psychologists within HCI are drawn to essentialist ontologies, and seek to achieve usability predominantly through consideration of user interface features. Specialists with a broader human-focus are mostly drawn to relational ontologies, and seek to understand how contextual factors interact with user interface features to shape experience and performance. Each ontology occupies ground within the HCI landscape. Both are now reviewed in turn. Usability evaluation methods are then briefly reviewed. While tensions between these two positions have dominated the evolution of usability in principle and practice, we can escape the impasse. A strategy for escaping longstanding tensions within usability will be presented, and future directions for usability within user experience frameworks will be indicated in the closing section.

## 15.3 LOCATING USABILITY WITHIN SOFTWARE: GUIDELINES, HEURISTICS, PATTERNS AND ISO 9126

### 15.3.1 Guidelines for Usable User Interfaces

Much early guidance on usability came from computer scientists such as Fred Hansen from Carnegie Mellon University (CMU) and Tony Wasserman, then at University of California, San Francisco (UCSF). Computer science has been strongly influenced by mathematics, where entities such as similar or equilateral triangles have eternal absolute intrinsic properties. Computer scientists seek to establish similar inherent properties for computer programs, including ones that ensure usability for interactive software. Thus initial guidelines on user interface design incorporated a technocentric belief that usability could be ensured via software and hardware features alone. A user interface would be inherently usable if it conformed to guidelines on, for example, naming, ordering and grouping of menu options, prompting for input types, input formats and value ranges for data entry fields, error message structure, response time, and undoing capabilities. The following four example guidelines are taken from Smith and Mosier's 1986 collection commissioned by the US Air Force (Smith and Mosier 1986):

| 1.0/4 + Fast Response |
|---|
| Ensure that the computer will acknowledge data entry actions rapidly, so that users are not slowed or paced by delays in computer response; for normal operation, delays in displayed feedback should not exceed 0.2 seconds. |
| **1.0/15 Keeping Data Items Short** |
| For coded data, numbers, etc., keep data entries short, so that the length of an individual item will not exceed 5-7 characters. |
| **1.0/16 + Partitioning Long Data Items** |
| When a long data item must be entered, it should be partitioned into shorter symbol groups for both entry and display. |
| **Example** |
| A 10-digit telephone number can be entered as three groups, NNN-NNN-NNNN. |
| **1.4/12 + Marking Required and Optional Data Fields** |
| In designing form displays, distinguish clearly and consistently between required and optional entry fields. |

Figure 15.9: Four example guidelines taken from Smith and Mosier's 1986 collection.

25 years after the publication of the above guidance, there are still many contemporary web site data entry forms whose users would benefit from adherence to these guidelines. Even so, while following guidelines can greatly improve software usability, it cannot guarantee it.

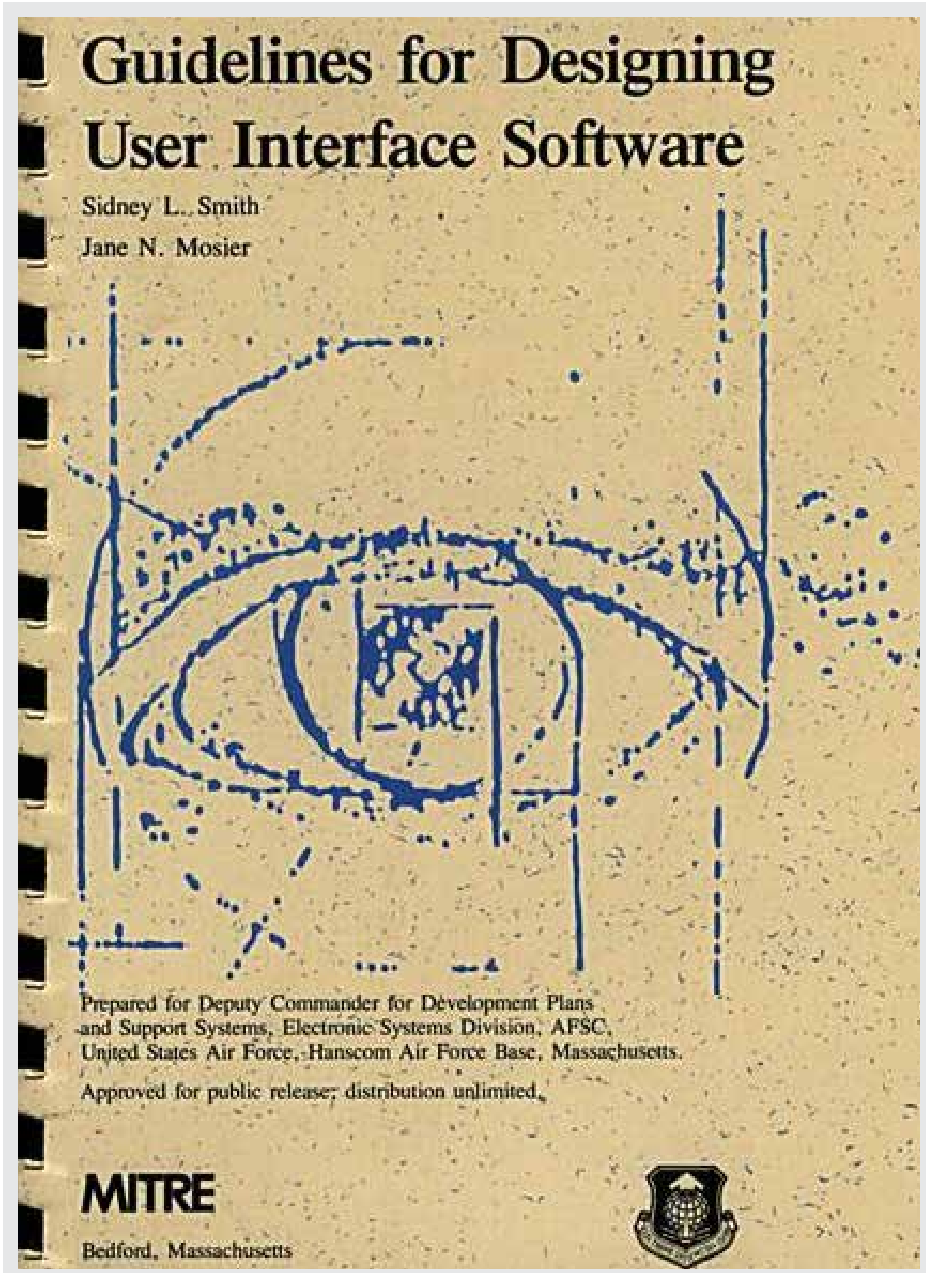



FIGURE 15.10: This Book Contains More Guidelines Than Anyone Could Imagine.

### 15.3.2 Manageable Guidance: Design Heuristics for Usable User Interfaces

My original paper copy of Smith and Mosier's guidelines occupies 10cm of valuable shelf space. It is over 25 years old and I have never read all of it. I most probably never will. There are simply too many guidelines there to make this worthwhile (in contrast, I have read complete style guides for Windows and Apple user interfaces in the past).

The bloat of guidelines collections did not remove the appeal of technocentric views of usability. Instead, hundreds of guidelines were distilled into ten heuristics by Rolf Mohlich and Jakob Nielsen. These were further assessed and refined into the final version of in *Heuristic Evaluation* (Nielsen 1994), an inspection method that examines software features for potential causes of poor usability. Heuristics generalise more detailed guidelines from collections such as Smith and Mosier. Many have a technocentric focus, e.g.:

**Visibility of system status**

The system should always keep users informed about what is going on, through appropriate feedback within reasonable time.

**User control and freedom**

Users often choose system functions by mistake and will need a clearly marked "emergency exit" to leave the unwanted state without having to go through an extended dialogue. Support undo and redo.

**Error prevention**

Even better than good error messages is a careful design which prevents a problem from occurring in the first place. Either eliminate error-prone conditions or check for them and present users with a confirmation option before they commit to the action.

**Recognition rather than recall**

Minimize the user's memory load by making objects, actions, and options visible. The user should not have to remember information from one part of the dialogue to another. Instructions for use of the system should be visible or easily retrievable whenever appropriate.

**Flexibility and efficiency of use**

Accelerators -- unseen by the novice user -- may often speed up the interaction for the expert user such that the system can cater to both inexperienced and experienced users. Allow users to tailor frequent actions.

Figure 15.11: Example heuristics orginally developed in Molich and Nielsen 1990 and Nielsen and Molich 1990.

Heuristic Evaluation became the most popular user-centred design approach in the 1990s, but has become less prominent with the move away from desktop applications. Quick and dirty user testing soon overtook Heuristic Evaluation (compare the survey of Venturi et al. 2006 with Rosenbaum et al. 2000).

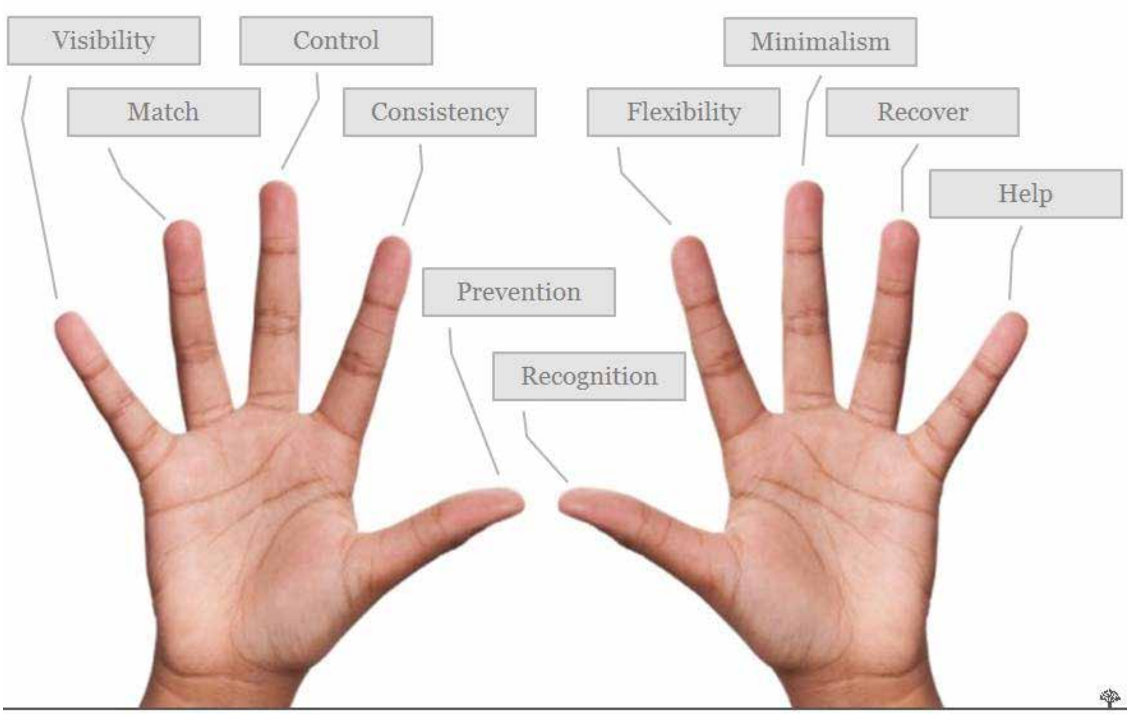



Figure 15.12: One Heuristic for Each Digit from Nielsen.

*Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.*

### 15.3.3 Invincible Intrinsics: Patterns and Standards Keep Usability Essential

Moves away from system-centric approaches within user-centred design have not signalled the end of usability methods that focus solely on software artefacts, with little or no attention to usage. This may be due to the separation of the usability communities (now user experience) from the software engineering profession. System-centredusability remains common in user interface pattern languages. For example, a pattern from Jenifer Tidwell updates Smith and Mosier style guidance for contemporary web designers (designinginterfaces.com/Input_Prompt).

**Pattern: Input Prompt**

Prefill a text field or dropdown with a prompt that tells the user what to do or type.

Figure 15.13: An example pattern from Jenifer Tidwell's Designing Interfaces.

The 1991 ISO 9126 standard on *Software Engineering – Product Quality* was strongly influenced by the essentialist preferences of computer science, with usability defined as:

> " a set of *[product]* attributes that bear on the effort needed for use, and on the individual assessment of such use, by a stated or implied set of users. "

This is the first of three definitions presented in this encyclopaedia entry. The attributes here are assumed to be software product attributes, rather than user interaction ones. However, the relational (contextual) view of usage favoured in HCI has gradually come to prevail. By 2001, ISO 9126 had been revised to define usability as:

> " (1') the capability of the software product to be understood, learned, used and attractive to the user, when used under specified conditions "

This revision remains product focused (essentialist), but the 'when' clause moved ISo 9126 away from a wholly essentialist position on usability by implicitly acknowledging the influence of a context of use ("specified conditions") that extends beyond "a stated or implied set of users".

In an attempt to align the technical standard ISO 9126 with the human factors standard ISO 9241 (see below), ISO 9126 was extended in 2004 by a fourth section on *quality in use*, resulting in an uneasy compromise between software engineers and human factors experts. This uneasy compromise persists, with the 2011 replacement standard for ISO 9126, ISO 25010 maintaining an essentialist view of usability. In ISO 25010, usability is both an intrinsic product quality characteristic *and* a subset of quality in use (comprising effectiveness, efficiency and satisfaction). As a product characteristic in ISO 25010, usability has the intrinsic subcharacteristics of:

- Appropriateness
- Recognisability
- Learnability
- Operability (degree to which a product or system has attributes that *make it* easy to operate and control - *emphasis added*)
- User error protection
- User interface aesthetics
- Accessibility

ISO 25010 thus had to include a note that exposed the internal conflict between software engineering and human factors world views:

> " Usability can either be specified or measured as a product quality characteristic in terms of its subcharacteristics, or specified or measured directly by measures that are a subset of quality in use. "

A similar note appears for learnability and accessibility. Within the world of software engineering standards, a mathematical world view clings hard to an essentialist position on usability. In HCI, where context has reigned for decades, this could feel incredibly perverse. However, despite HCI's multi-factorial understanding of usability, which follows automatically from a contextual position, HCI evangelists' anger over poor usability always focuses on software products. Even though users, tasks and contexts are all known to influence usability, only hardware or software should be changed to improve usability, endorsing the software engineers' position within ISO 25010 (attributes *make software easy to operate and control*). Although HCI's world view typically rejects essentialist monocausal explanations of usability, when getting angry on the user's behalf, the software always gets the blame.

It should be clear that issues here are easy to state but harder to unravel. The stalemate in ISO 25010 indicates a need within HCI to give more weight to the influence of software design on usability. If users, tasks and contexts must not be changed, then the only thing that we can change is hardware and/or software. Despite the psychological marginalisation of designers' experience and expertise when expressed in guidelines, patterns and heuristics, these can be our most critical resource for achieving usability best practice. We should bear this in mind as we move to consider HCI's dominant contextual position on usability.

## 15.4 LOCATING USABILITY WITHIN INTERACTION: CONTEXTS OF USE AND ISO STANDARDS

The tensions within international standards could be seen within Nielsen's Heuristics, over a decade before the 2004 ISO 9126 compromise. While the five sample heuristics in the previous section focus on software attributes, one heuristic focuses on the relationship between a design and its context of use (Nielsen 1994):

> “ **Match between system and the real world**
> The system should speak the users' language, with words, phrases and concepts familiar to the user, rather than system-oriented terms. Follow real-world conventions, making information appear in a natural and logical order. ”

This relating of usability to the ‘real world’ was given more structure in the ISO 9241-11 Human Factors standard, which related usability to the usage context as the:

> “ Extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use ”

This is the second of three definitions presented in this encyclopaedia entry. Unlike the initial and revised ISO 9126 definitions, it was not written by software engineers, but by human factors experts with backgrounds in ergonomics, psychology and similar.

ISO 9241-11 distinguishes three component factors of *usability*: effectiveness, efficiency, satisfaction. These result from multi-factorial interactions between users, goals, contexts and a software product. Usability is not a characteristic, property or quality, but an *extent* within a multi-dimensional space. This extent is evidenced by what people can actually achieve with a software product and the costs of these achievements. In practical terms, any judgement of usability is a holistic assessment that combines multi-faceted qualities into a single judgement.

Such single judgements have limited use. For practical purposes, it is more useful to focus on separate specific qualities of user experience, i.e., the extent to which *thresholds* are met for different qualities. For example, a software product may not be deemed usable if key tasks cannot be performed in normal operating contexts within an acceptable time. Here, the focus would be on *efficiency* criteria. There are many usage contexts where time is limited. The bases for time limits vary considerably, and include physics (ballistics in military combat), physiology (medical trauma), chemistry (process control) or social contracts (newsroom print/broadcast deadlines).

*Effectiveness* criteria add to the complexity of quality thresholds. A military system may be efficient, but it is not effective if its use results in what is euphemistically called 'collateral damage', including 'friendly fire' errors. We can imagine trauma resuscitation software that enables timely responses, but leads to avoidable 'complications' (another domain euphemism) after a patient has been stabilised. A process control system may support timely interventions, but may result in waste or environmental damage that limits the effectiveness of operators' responses. Similarly, a newsroom system may support rapid preparation of content, but could obstruct the delivery of high quality copy.

For *satisfaction*, usage could be both objectively efficient and effective, but cause uncomfortable user experiences that give rise to high rates of staff turnover (as common in call centres). Similarly, employees may thoroughly enjoy a fancy multimedia fire safety training title, but it could be far less effective (and thus potentially deadly) compared to the effectiveness of a boring instructional text–with-pictures version.

ISO 9241-11's three factors of usability have recently become five in by ISO 25010's quality in use factors:

- Effectiveness
- Efficiency
- Satisfaction
- Freedom from risk
- Context coverage

The two additional factors are interesting. *Context coverage* is a broader concept than the contextual fit of the *match between system and* Nielsen's *Match between System and Real World* heuristic (Nielsen 1994). It extends *specified users* and *specified goals* to potentially any aspect of a context of use. This should include all factors relevant to *freedom from risk*, so it is interesting to see this given special attention, rather than trusting *effectiveness* and *satisfaction* to do the work here. However, such piecemeal extensions within ISO 25010 open up the question of comprehensiveness and emphasis. For example, why are factors such as ease of learning either overlooked or hidden inside efficiency or effectiveness?

FIGURE 15.14: ISO Accessibility Standard Discussion.



### 15.4.1 Contextual Coverage Brings Complex Design Agendas

Relational positions on usability are inherently more complex than essentialist ones. The latter let, interactive systems be *inspected* to assess their usability potential on the basis of their design features. Essentialist approaches remain attractive because evaluations can be fully resourced through guidelines, patterns and similar expressions of best practice for interaction design. Relational approaches require a more complex set of co-ordinated methods. As relational positions become more complex, as in the move from ISO 9241-11 to ISO 25010, a broader range of evaluation methods is required. Within the relational view, usability is the result of a set of complex interactions that manifests itself in a range of usage factors. It is very difficult to see how a single evaluation method could address all these factors. Whether or not this is possible, no such method currently exists.

Relational approaches to usability require a range of evaluation methods to establish its *extent*. Extent introduces further complexities, since all identified usability factors must be *measured*, then judgements must be made as to whether achieved extents are adequate. Here, usability evaluation is not a simple matter of inspection, but instead it becomes a complex logistical operation focused on implementing a *design agenda*.

An *agenda* is list of things to be done. A *design* agenda is therefore a list of design tasks, which need to be managed within an embracing development process. There is an implicit design agenda in ISO 9241-11, which requires interaction designers to identify target beneficiaries, uasge goals, and levels of efficiency, effectiveness and satisfaction for a specific project. Only then is detailed robust usability evaluation possible. Note that this holds for ISO 9241-11 and similar evaluation philosophies. It does not hold for some other design philosophies (e.g., Sengers and Gaver 2006) that give rise to different design agendas.

A key task on the ISO 9241-11 evaluation agenda is thus measuring the extent of usability through a co-ordinated set of metrics, which will typically mix quantitative and qualitative measures, often with a strong bias towards one or the other. However, measures only enable evaluation. To evaluate, measures need to be accompanied by targets. Setting such targets is another key task from the ISO 9241-11 evaluation agenda. This is often achieved through the use of generic *severity scales*. To use such resources, evaluators need to interpret them in specific project contexts. This indicates that re-usable evaluation resources are not complete re-usable solutions. Work is required to turn these resources into actionable evaluation tasks.

For example, the two most serious levels of Chauncey Wilson's problem severity scale (Wilson 1999) are:

- Level 1 — Catastrophic error causing irrevocable loss of data or damage to the hardware or software. The problem could result in large-scale failures that prevent many people from doing their work. Performance is so bad that the system cannot accomplish business goals.

- Level 2 — Severe problem causing possible loss of data. User has no workaround to the problem. Performance is so poor that … universally regarded as 'pitiful'.

Each severity level requires answers to questions about specific measures and contextual information, i.e., how should the following be interpreted in a specific project context: 'many prevented from doing work'; 'cannot accomplish business goals'; 'performance regarded as pitiful'. These top two levels also require information about the software product: 'loss of data'; 'damage to hardware of software'; 'no workaround'.

Wilson's three further lower level scales add considerations such as: 'wasted time', 'increased error or learning rates', and 'important feature not working as expected'. These all set a design agenda of questions that must be answered. Thus to know that performance is regarded as pitiful, we would need to choose to measure relevant subjective judgments. Other criteria are more challenging, e.g., how would we know whether time is wasted, or whether business goals cannot be accomplished? The first depends on values. The idea of 'wasting' time (like money) is specific to some cultural contexts, and also depends on how long tasks are expected to take with a new system, and how much time can be spent on learning and exploring. As for business goals, a business may seek, for example, to be seen as socially and environmentally responsible, but may not expect every feature of every corporate system to support these goals.

Once thresholds for severity criteria have been specified, it is not clear how designers can or should trade off factors such as efficiency, effectiveness and satisfaction against each other. For example, users may not be satisfied even when they exceed target efficiency and effectiveness, or conversely they could be satisfied even when their performance should not warrant that relative to design targets. Target levels thus guide rather than dictate the interpretation of results and how to respond to them.

Method requirements thus differ significantly between essentialist and relational approaches to usability. For high quality evaluation based on any relational position, not just ISO 9241-11's, evaluators must be able to modify and combine existing re-usable resources for specific project contexts. Ideally, the re-usable resources would do most of the work here, resulting in efficient, effective and satisfying usability evaluation. If this is not the case, then high quality usability evaluation will present complex logistical challenges that require extensive evaluator expertise and project specific resources.

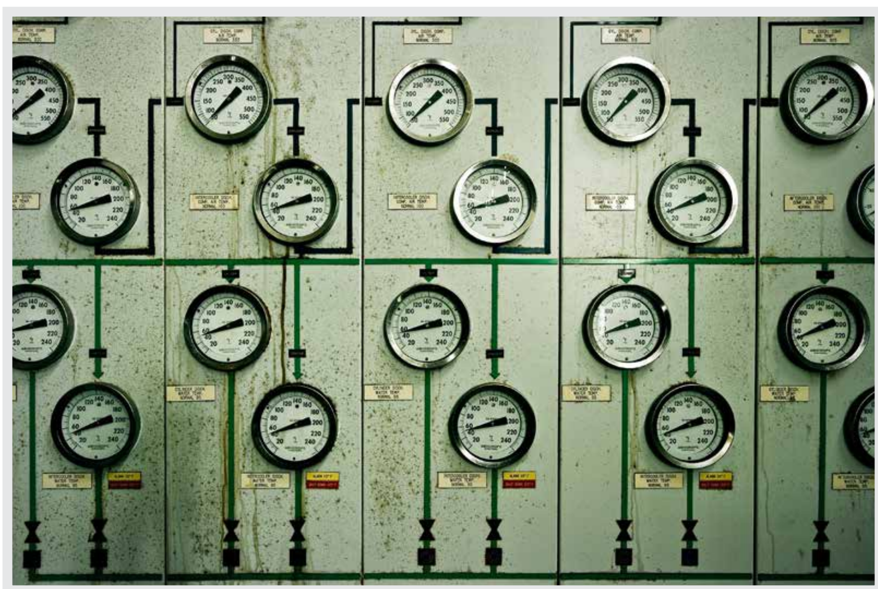


Figure 15.15: Relational Approaches to Usability Require Multiple Measures.



## 15.5 THE DEVELOPMENT OF USABILITY EVALUATION: TESTING, MODELLING AND INSPECTION

Usability is a contested historical term that is difficult to replace. User experience specialists have to refer to usability, since it is a strongly established concept within the IT landscape. However, we need to exercise caution in our use of what is essentially a flawed concept. Software is not usable. Instead, *software gets used*, and the resulting user experiences are a composite of several qualities that are shaped by product attributes, user attributes and the wider context of use.

Now, squabbles over definitions will not necessarily impact practice in the 'real world'. It is possible for common sense to prevail and find workarounds for what could well be semantic distractions with no practical import. However, when we examine usability evaluation methods, we do see that different conceptualisations of usability result in differences over the *causes* of good and poor usability.

Essentialist usability is, causally *homogeneous*. This means that all causes of user performance are of the same type, i.e., due to technology. System-centred inspection methods can identify such causes.

Contextual usability is causally *heterogeneous*. This means that causes of user performance are of different types, some due to technologies, others due to some aspect(s) of usage contexts, but most due to interactions between both. Several evaluation and other methods may be needed to identify and relate a nexus of causes.

Neither usability paradigm (i.e., essentialist or relational) has resolved the question of relevant *effects*, i.e., what counts as evidence of good or poor usability, and thus there are few adequate methods here. Essentialist usability can pay scant attention to effects (Lavery et al. 1997): who cares what poor design will do to users, it's bad enough that it's poor design! Contextual usability has more focus on effects, but there is limited consensus on the sort of effects that should count as evidence of poor usability. There are many examples of what could count as evidence, but what actually should is left to a design team's judgement.

Some methods can predict effects. The *GOMS* model (Goals, Operators, Methods, and Selection rules) predicts effects on expert error free task completion time, which is useful in some project contexts (Card et al 1980, John and Kieras 1996). For example, external processes may require a task to be completed within a maximum time period. If predicted expert error free task completion time exceeds this, then it is highly probable that non-expert error prone task completion take even longer. Where interactive devices such as in-car systems distract attention from the main task (e.g., driving), then time predictions are vital. Recent developments such as CogTool (Bellamy et al. 2011) have given a new lease of life to practical model-based evaluation in HCI. More powerful models than GOMS are now being integrated into evaluation tools (e.g., Salvucci 2009).


Figure 15.16: Model-Based methods can predict how long drivers could be distracted, and much more.



Usability work can thus be expected to involve a mix of methods. The mix can be guided by high level distinctions between methods. Evaluation methods can be *analytical* (based on examination of an interactive system and/or potential interactions with it) or *empirical* (based on actual usage data). Some analytical methods require the construction of one or more models. For example, GOMS models the relationships between software and human performance. Software attributes in GOMS all relate to user input methods at increasing levels of abstraction from the keystroke level up to abstract command constructs. System and user actions are interleaved in task models to predict users' methods (and execution times at a keystroke level of analysis).

### 15.5.1 Analytical and Empirical Evaluation Methods, and How to Mix Them

*Analytical evaluation methods* may be *system-centred* (e.g., Heuristic Evaluation) or *interaction-centred* (e.g., Cognitive Walkthrough). Design teams use the resources provided by a method (e.g., heuristics) to identify strong and weak elements of a design from a usability perspective. Inspection methods tend to focus on the *causes* of good or poor usability. *System-centred* inspection methods focus solely on software and hardware features for attributes that will promote or obstruct usability. *Interaction-centred* methods focus on two or more causal factors (i.e., software features, user characteristics, task demands, other contextual factors).

*Empirical evaluation methods* focus on evidence of good or poor usability, i.e., the positive or negative *effects* of attributes of software, hardware, user capabilities and usage environments. User testing is the main project-focused method. It uses project-specific resources such as test tasks, users, and also measuring instruments to expose usability problems that can arise in use. Also, essentialist usability can use empirical experiments to demonstrate superior usability arising from user interface components (e.g., text entry on mobile phones) or to optimise tuning parameters (e.g., timings of animations for windows opening and closing). Such experiments assume that the test tasks, test users and test contexts allow generalisation to other users, tasks and contexts. Such assumptions are readily broken, e.g., when users are very young or elderly, or have impaired movement or perception.

Analytical and empirical methods emerged in rapid succession, with empirical methods emerging first in the 1970s as simplified psychology experiments (for examples, see early volumes of International Journal of Man-Machine Studies 1969-79). Model-based approaches followed in the 1980s, but the most practical ones are all variants of the initial GOMS method (John and Kieras 1996). Model-free inspection methods appeared at the end of the 1980s, with rapid evolution in the early 1990s. Such methods sought to reduce the cost of usability evaluation by *discounting* across a range of resources, especially users (none required, unlike user testing), expertise (transferred by heuristics/models to novices) or extensive models (none required, unlike GOMS).


Figure 15.17: Chicken Fajitas Kit: everything you need except chicken, onion, peppers, oil, knives, chopping board, frying pan, stove etc. Usability Evaluation Methods are very similar - everything is fine once you've worked to provide everything that's missing.

Achieving balance in a mix of evaluation methods is not straightforward, and requires more than simply combining analytical and empirical methods. This is because there is more to usability work than simply choosing and using methods. Evaluation methods are as complete as a Chicken Fajita Kit, which contains very little of what is actually needed to make Chicken Fajitas: no chicken, no onion, no peppers, no cooking oil, no knives for peeling/coring and slicing, no chopping board, no frying pan, no stoves etc. Similarly, user testing 'methods' as published miss out equally vital ingredients and project specific resources such as participant recruitment criteria, screening questionnaires, consent forms, test task selection criteria, test (de)briefing scripts, target thresholds, and even data collection instruments, evaluation measures, data collation formats, data analysis methods, or reporting formats. There is no complete published user testing method that novices can pick up and use 'as is'. All user testing requires extensive project-specific planning and implementation. Instead, much usability work is about configuring and combining methods for project-specific use.

### 15.5.2 The Only Methods are the Ones that You Complete Yourselves

When planning usability work, it is important to recognise that so-called 'methods' are more strictly loose collections of resources better understood as 'approaches'. There is much work in getting usability work to work, and as with all knowledge-based work, methods cannot copied from books and applied without a strong understanding of fundamental underlying concepts. One key consequence here is that only specific instances of methods can be compared in empirical studies, and thus credible research studies cannot be designed to collect evidence of systematic reliable differences between different usability evaluation methods. All methods have unique usage settings that require project-specific resources, e.g., for user testing, these include participant recruitment, test procedures and (de-)briefings. More generic resources such as problem extraction methods (Cockton and Lavery 1999) may also vary across user testing contexts. These inevitably obstruct reliable comparisons.


Figure 15.18 A-B: Dogs or Cats: which is the better pet? It all depends on what sorts of cats and dogs you compare, and how you compare them. The same is true of evaluation methods.

Consider a simple comparison of heuristic evaluation against user testing. Significant effort would be required to allow a fair comparison. For example, if the user testing asked test users to carry out fixed tasks, then heuristic evaluators would need to explore the same system using the same tasks. Any differences and similarities between evaluation results for the two methods would not generalise beyond these fixed tasks, and there are also likely to be extensive evaluation effects arising from individual differences in evaluator expertise and performance. If tasks are not specified for the evaluations, then it will not be clear whether differences and similarities between results are due to the approaches used or to the unrecorded tasks within for the evaluations. Given the range of resources that need to be configured for a specific user test, it is simply not possible to control all known potential confounds (still less all currently unknown ones). Without such controls, the main sources of differences between methods may be factors with no bearing on actual usability.

The tasks carried out by users (in user testing) or used by evaluators (in inspections or model specifications) are thus one possible confound when comparing evaluation approaches. So too are evaluation measures and target thresholds. Time on task is a convenient measure for usability, and for some usage contexts it is possible to specify worthwhile targets, e.g., for supermarket checkouts thetarget time to check out representative trolleys of purchases could be 30 minutes for 10 typical trolley loads of shopping). However, in many settings, there are no time thresholds for efficient use that can be used reliably (e.g., time to draft and print a one page business letter, as opposed to typing one in from a paper draft or a dictation).

Problems associated with setting thresholds are compounded by problems associated with choosing the measures for which thresholds are required. A wide range of potential measures can be chosen for user testing. For example, in 1988, usability specialists from Digital Equipment Corporation and IBM (Whiteside et al. 1988) published a long list of possible evaluation measures, including:

Measure without measure: there's so much scope for scoring

- Counts of:
  - commands used
  - repetitions of failed commands
  - runs of successes and of failures
  - good and bad features recalled by users
  - available commands not invoked/regressive behaviours
  - users preferring your system
- Percentage of tasks completed in time period
- Counts or percentages of:
  - errors
  - superior competitor products on a measure
- Ratios of
  - successes to failures
  - favourable to unfavourable comments
- Times
  - to complete a task
  - spent in errors
  - spent using help or documentation
- Frequencies
  - of help and documentation use
  - of interfaces misleading users

- users needing to work around a problem
- users disrupted from a work task
- users losing control of the system
- users expressing frustration or satisfaction

No claims were made for the comprehensiveness of the full list of measures that were known to have been used up to the point of publication within Digital Equipment Corporation or IBM. What was clear was a position that project teams must choose their own metrics and thresholds. No methods yet exist to reliably support such choices.

There are no universal measures of usability that are relevant to every software development project. Interestingly, Whiteside et al. (1988) was the publication that first introduced *contextual design* to the HCI community. Its main message was that existing user testing practices were delivering far less value for design than contextual research. A hope was expressed that once contexts of use were better understood, and contextual insights could be shown to inform successful design across a diverse range of projects, then new contextual measures would be found for more appropriate evaluation of user experiences. Two decades elapsed before bases for realising this hope emerged within HCI research and professional practice. The final sections of this encyclopaedia entry explore possible ways forward.

### 15.5.3 Sorry to Disappoint You But ...

To sum up the position so far:

1. There are fundamental differences on the nature of usability, i.e., it is either an inherent property of interactive systems, or an emergent property of usage. There is no single definitive answer to what usability ‘is’.

Usability is only an inherent measurable property of all interactive digital technologies for those who refuse to think of it in any other way.

2. There are no universal measures of usability, and no fixed thresholds above or below which all interactive systems are or are not usable. There are no *universal*, robust, objective and reliable metrics. There are no evaluation methods that unequivocally determine whether or not an interactive system or device is usable, or to what extent. All positions here involve hard won expertise, judgement calls, and project-specific resources beyond what all documented evaluation methods provide.

3. Usability work is too complex and project-specific to admit generalisable methods. What are called ‘methods’ are more realistically ‘approaches’ that provide loose sets of resources that need to be adapted and configured on a project by project basis. There are no reliable pre-formed methods for assessing usability. Each method in use is unique, and relies heavily on the skills and knowledge of evaluators, as well as on project-specific resources. There are no off-the-shelf evaluation methods. Evaluation methods and metrics are not completely documented in any literature. Developing expertise in usability measurement and evaluation requires far more than reading about methods, learning how to apply them, and through this alone, becoming proficient in determining whether or not an interactive system or device is usable, and if so, to what extent. Even system-centred essentialist methods leave gaps for evaluators to fill (Cockton et al. 2004, Cockton et al. 2012).

The above should be compared with the four opening propositions, which together constitute an attractive ideology that promises certainties regardless of evaluator experience and competence. Each proposition is not wholly true, and can be mostly false. Evaluation can never be an add-on to software development projects. Instead, the scope of usability work, and the methods used, need to be planned with other design and development activities. Usability evaluation requires supporting resources that are an integral part of every project, and must be developed there.

The tension between essentialist and relational conceptualisations of usability is only the tip of the iceberg of challenges for usability work. Not only is it not clear what usability is (although competing definitions are available), but it is also not clear specifically how usability should be assessed outside of the contexts of specific projects. What matters in one context may not matter in another. Project teams must decide what matters. The usability literature can indicate possible measure of usability, but none are universally applicable. The realities of usability work are that each project brings unique challenges that require experience and expertise to meet them. Novice evaluators cannot simply research, select and apply usability evaluation methods. Instead, actual methods in use are the critical *achievement* of all usability work.

Methods are *made* on the ground on a project by project basis. They are not archived 'to go' in the academic or professional literature. Instead there are two starting points. Firstly, there are literatures on a range of approaches that provide some re-usable resources for evaluators, but require additional information and judgement within project contexts before practical methods can be completed. Secondly, there are detailed case studies of usability work within specific projects. Here the challenge for evaluators is to identify resources and practices within the case study that would have a good fit with other project contexts, e.g., a participant recruitment procedure from a user testing case study may be re-usable in other projects, perhaps with some modifications.

Readers could reasonably draw the conclusion from the above that usability is an attractive idea in principle that has limited substance in reality. However, the reality is that we all continue to experience frustrations when using interactive digital technologies, and often we would say that we do find them difficult to use. Even so, frustrating user experiences may not be due to some single abstract construct called ‘usability’, but instead be the result of unique complex interactions between people, technology and usage contexts. Interacting factors here must be considered together. It is not possible to form judgements on the severity of isolated usage difficulties, user discomfort or dissatisfaction. Overall judgements on the quality of interactive software must balance what can be achieved through using it against the costs of this use. There are no successful digital technologies without what could be usability flaws to some HCI experts (I can *always* find some!). Some technologies appear to have severe flaws, and are yet highly successful for many users. Understanding why this is the case provides insights that move us away from a primary focus on usability in interaction design.

## 15.6 WORTHWHILE USABILITY: WHEN AND WHY USABILITY MATTERS, AND HOW MUCH

While writing the previous section, I sought advice via Facebook on transferring contacts from my vintage Nokia N96 mobile phone to my new iPhone. One piece of advice turned out to be specific to Apple computers, but was still half-correct for a wintel PC. Eventually, I identified a possible data path that required installing the Nokia PC suite on my current laptop, backing up contacts from my old phone to my laptop, downloading a freeware program that would convert contacts from Nokia’s proprietary backup format into a text format for spreadsheets/databases (comma separated values - .csv), failing to import it into a cloud service, importing it into the Windows Address Book on my laptop (after spreadsheet editing), and then finally synchronising the contacts instead via iTunes with my new iPhone.

### 15.6.1 A Very Low Frequency Multi-device Everyday Usability Story

From start to finish, my phone number transfer task took two and a half hours. Less than half of my contacts were successfully transferred, and due to problems in the spreadsheet editing, I had to transfer contacts in a form that required further editing on my iPhone or in the Windows contacts folder.

Focusing on the ISO 9241-11 definition of usability, what can we say here about the usability of a complex ad hoc overarching product-service system involving social networks, cloud computing resources, web searches, two component product-service systems (Nokia 96 + Nokia PC Suite, iPhone + iTunes) and Windows laptop utilities?

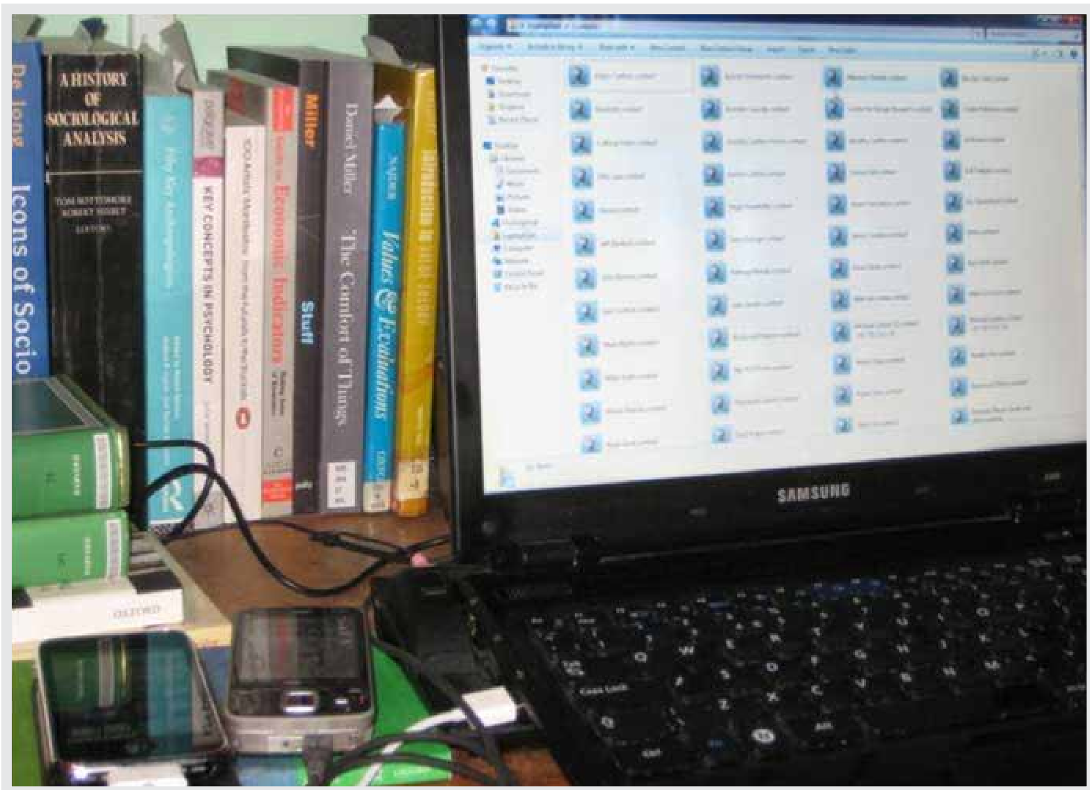



Figure 15.19: A Tale of Two Mobiles and Several Software Utilities.



Was it efficient taking 2.5 hours over this? Around 30 minutes each were spent on:

1. web searches, reading about possible solutions, and a freeware download
2. installing mobile phone software (new laptop since I bought the Nokia), attempts to connect Nokia to laptop, laptop restart, successful backup, extraction to .csv format with freeware

3. exploring a cloud email contacts service, failing to upload to it.
4. test upload to Windows address book, edits to improve imports, failed edits of phone numbers, successful import
5. Synchronisation of iPhone and iTunes

To reach a judgement on efficiency, we need to first bear in mind that during periods of waiting (uploads, downloads, synchronisations, installations), I proof read my current draft of this entry and corrected it. This would have taken 30 minutes anyway. Secondly, I found useful information from the web searches that lead me to the final solution. Thirdly, I had to learn how to use the iTunes synchronisation capabilities for iPhones, which took around 10 minutes and was an essential investment for the future. However, I wasted at least 30 minutes on a cloud computing option suggested on Facebook (I had to add email to the cloud service before failing to upload from the .csv file). There were clear usability issues here, as the email service gave no feedback as to why it was failing to upload the contacts. There is no excuse for such poor interaction design in 2011, which forced me to try a few times before I realised that it would not work, at least with the data that I had. Also, the extracted phone numbers had text prefixes, but global search and replace in the spreadsheet resulted in data format problems that I could not overcome. I regard both of these as usability problems, one due to the format of the telephone numbers as extracted, and one due to the bizarre behaviour of a well known spreadsheet programme.

I have still not answered the question of whether this was efficient in ISO 9241-11 terms. I have to conclude that it was not, but this was partly due to my lack of knowledge on co-ordinating a complex combination of devices and software utilities. However, back when contacts were only held on mobile phone SIMs, the transfer would have taken a few minutes in a phone store. So, current usability problems here are due to the move to storing contacts in more comprehensive formats separately from a mobile phone's SIM. However, while there used to be a more efficient option, most of us now make use of more comprehensive phone memory contacts, and thus the previous fast option was at the cost of the most primitive contact format imaginable. So while the activity was relatively inefficient, there are potentially compensating reasons for this.

The only genuine usability problems relate to the lack of feedback in the cloud-based email facility, the extracted phone number formats, and bizarre spreadsheet behaviour. However, I only explored the cloud email option following advice via Facebook. My experience of problems here was highly contextual. For the other two problems, if the second problem had not existed, then I would never have experienced the third.

There are clear issues of efficiency. At best this took twice as long as it should have once interleaved work and much valuable re-usable learning are discounted. However, the causes of this inefficiency are hard to pin-point within the complex socially shaped context within which I was working.

Effectiveness is easy to evaluate. I only transferred just under 50% of the contacts. Note how straightforward the analysis is here when compared to efficiency in relation to a complex product-service system.

On balance, you may be surprised to read that I was fairly satisfied. Almost 50% is better than nothing. I learned how to synchronise my iPhone via iTunes for the first time. I made good use of the waits in editing this encyclopaedia entry. I was not in any way *happy* though, and I remain dissatisfied over the phone number formats, inscrutable spreadsheet behaviour and mute import facility on a top three free email facility.

### 15.6.2 And the Moral of My Story Is: It was Worth It, on Balance

What overall judgement can we come to here? On a *binary* basis, the final data path that I chose *was usable*. An abandoned path was not, so I did encounter one unusable component during my attempt to transfer phone numbers. As regards a more realistic *extent* of usability (as opposed to binary usable vs. unusable), we must trade off factors such as efficiency, effectiveness and satisfaction against each other. I could rate the final data path as 60% usable, with effective valuable learning counteracting the ineffective loss of over half of my contacts, which I had to subsequently enter manually. I could raise substantially this to 150% by adding the value of the resulting example for this encyclopaedia entry! It reveals the complexity of evaluating usability of interactions involving multiple devices and utilities. Describing usage is straightforward: judging its quality is not.

So, poor usability is still with us, but it tends to arise most often when we attempt to co-ordinate multiple digital devices across a composite ad-hoc product-service system. Forlizzi (2008) and others refer to these now typical usage contexts as *product ecologies*, although some (e.g., Harper et al. 2008) prefer the term product *ecosystems*, or product-service ecosystems (ecology is the discipline of ecosystems, not the systems themselves).

Components that are usable enough in isolation are less usable in combination. Essentialist positions on usability become totally untenable here, as the phone formats can blame the bizarre spreadsheet and vice-versa. The effects of poor usability are clear, but the causes are not. Ultimately, the extent of usability, and its causes in such settings, is a matter of interpretation based on judgements of the value achieved and the costs incurred.

Far from being an impasse, regarding usability as *a matter of interpretation* actually opens up a way forward for evaluating user experiences. It is possible to have robust interpretations of efficiency, effectiveness and satisfaction, and robust bases for overall assessments of how these trade-off against each other. To many, these bases will appear to be subjective, but this is not a problem, or at least it is far less of a problem than acting erroneously as if we have generic universal objective criteria for the existence or extent of usability in any interactive system. To continue any quest for such criteria is most definitely inefficient and ineffective, even if the associated loyalties to seventeenth century scientific values bring some measure of personal (subjective) satisfaction.

It is poor usability that focused HCI attention in the 1980s. There was no positive conception of *good* usability. Poor usability could degrade or even destroy the intended value of an interactive system. However, good usability can not *donate* value beyond that intended by a design team. Usability evaluation methods are focused on finding problems, not on finding successes (with the exception of Cognitive Walkthrough). Still, experienced usability practitioners know that an evaluation report should begin by commending the strong points of a design, but these are not what usability methods are optimised to detect.

Realistic relevant evaluations must assess *incurred costs* relative to *achieved benefits*. When transferring my contacts between phones, I experienced the following problems and associated costs:

1. Could not upload contacts into cloud email system, despite several attempts (cost: wasted 30 minutes)

2. Could not understand why I could not upload contacts into cloud email system (costs: prolonged frustration, annoyance, mild anger, abusing colleagues' company #1)

3. Could not initiate data transfer from Nokia phone first time, requiring experiments and laptop restart as advised by Nokia diagnostics (cost: wasted 15 minutes)

4. Over half of my contacts did not transfer (future cost: 30-60 further minutes entering numbers, depending on use of laptop or iPhone, in addition to 15 minutes already spent finding and noting missing contacts)

5. Deleting type prefixes (e.g., TEL CELL) from phone numbers in a spreadsheet resulted in an irreversible conversion to a scientific format number (cost: 10 wasted minutes, plus future cost of 30-60 further minutes editing numbers in my phone, bewilderment, annoyance, mild anger, abusing colleagues' company #2)

6. Had to set a wide range of synchronisation settings to restrict synchronisation to contacts (cost extra 10 minutes, initial disappointment and anxiety)

7. Being unable to blame Windows for anything (this time)!

By forming the list above, I have taken a position on what, in part, would count as poor usability. To form a judgement as to whether these costs were *worthwhile*, I also need to take a position on positive outcomes and experiences:

1. an opportunity to ask for, and receive, help from Facebook friends (realising some value of existing social capital)

2. a new email address gilbertcockton@... via an existing cloud computing account (future value unknown at time, but has since proved useful)

3. Discovered a semi-effective data path that transferred almost half of my contacts to my iPhone (saved: 30-60 minutes of manual entry, potential re-usable knowledge for future problem solving)

4. Learned about a nasty spreadsheet behaviour that could cause problems in the future unless I find out how to avoid it (future value potentially zero)

5. Learned about the Windows address book and how to upload new contacts as .csv files (very high future value – at the very least PC edits/updates are faster than iPhone, with very easy copy/paste from web and email)

6. Learned how to synchronise my new iPhone with my laptop via iTunes (extensive indubitable future value, repeatedly realised during the editing of this entry, including effortless extension to my recent new iPad)

7. Time to proof the previous draft of this entry and edit the next version (30 minutes of effective work during installs, restarts and uploads)

8. Sourced the main detailed example for this encyclopaedia entry (hopefully as valuable to you as a reader as to me as a writer:I've found it really helpful)

In many ways the question as to whether the combined devices and utilities were 'usable' has little value, as does any question about the extent of their combined usability. A more helpful question is whether the interaction was *worthwhile*, i.e., did the achieved resulting benefits justify the expended costs? *Worth* is a very useful English word that captures the relationship between costs and benefits: achieved *benefits* are (not) *worth* the incurred costs. Worth relates positive value to negative value, considering the balance of both, rather than, as in the case of poor usability, mostly or wholly focusing on negative factors.

Figure 15.20: Usability Verdict: Not Guilty.

So, did my resulting benefits justify my expended costs? My answer is yes, which is why I was satisfied at the time, and am more satisfied now as frustrations fade and potential future value has been steadily realised. Given the two or three usability problems encountered, and their associated costs, it is quite clear that the interaction could have been more worthwhile (increased value at lower costs), but this position is more clear cut than having to decide on the extent and severity of usability problems in isolation. The interaction would have been more worthwhile in the absence of usability problems (but I would not have this example). It would have also been more worthwhile if I'd already known in advance how to extract contacts from a Nokia backup file in a format where they could have been uploaded into the Windows address book of contacts. Still better, the utility suite that came with my phone could have had.cvs file import/export. Perhaps the best solution would be for phones to enable Windows to import contacts from them. Also, if I had used my previous laptop, the required phone utility suite was already installed and there should have been no initial connection problems. There were thus ways of reducing costs and increasing value that would not involve modifications to the software that I used, but would instead have replaced them all with *one simple purpose built tool*. None of the experienced usability problems would have been fixed. Once the complexity of the required data path is understood, it is clear that the best thing to do is to completely re-engineer it. Obliteration beats iteration here.

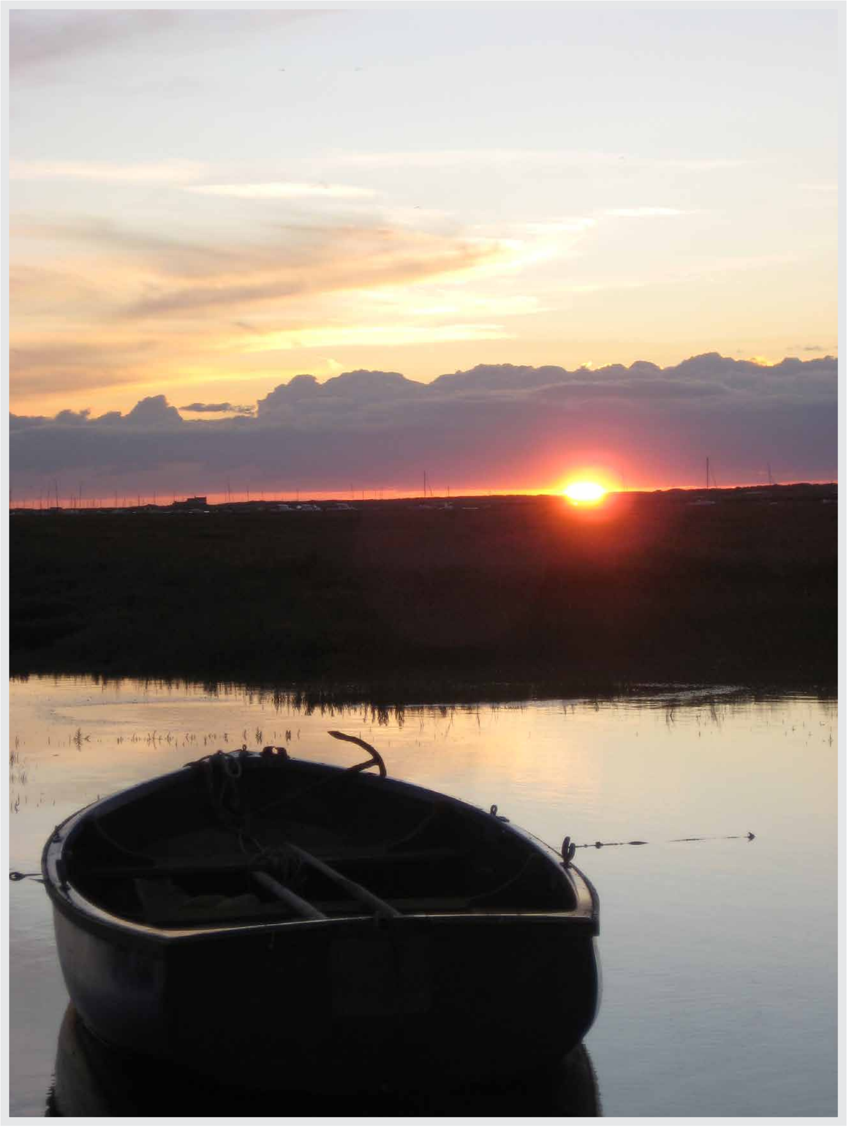


FIGURE 15.21: At the end of the day, you have to look at the big picture.

*Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.*

## 15.6.3 Usability is Only One Part of a BIG Interaction Design Picture

By considering usability within the broader context of experience and outcomes, many dilemmas associated with usability in isolation disappear. This generalises to human-centred design as a whole. In his book *Change by Design*, Tim Brown, CEO of IDEO, builds a compelling case for the human-centred practices of multi-disciplinary design teams. Even so, he acknowledges the lack of truly compelling stories that fully establish the importance of human-centred design to innovation, since these are undermined by examples of people regularly surmounting inconveniences (Brown 2009, pp.39-40), to which I have just added above. Through examples such as chaining bicycles to park benches, Brown illustrates *worth in action*: the benefit (security of bike) warrants the cost (finding a nearby suitable fixed structure to chain to). The problem with usability evaluations is that they typically focus on incurred costs without a balancing focus on achieved benefits. Brown returns to the issue of balance in his closing chapter, where design thinking is argued to achieve balance through its integrative nature (p.229).

Human-centred contributions to designs are just one set of inputs. Design success depends on effective integration of all its inputs. Outstanding design always overachieves, giving users/owners/sponsors far more than they were expecting. The best design is thus **B**alanced, **I**ntegrative and **G**enerous – or plain **BIG** for short. Usability needs to fit into the big picture here.

Usability evaluation finds usage problems. These need to be understood holistically in the full design context before possible solutions can be proposed. Usability evaluation cannot function in isolation, at least, not without isolating the usability function. Since the early 90s, usability specialists have had a range of approaches to draw on, which, once properly adapted, configured and combined can provide highly valuable inputs to the iterative development of interaction designs. Yet we continue to experience interaction design flaws, such as lack of instructive actionable feedback on errors and problems, which can and should be eliminated. However, appropriate use of usability expertise is only one part of the answer. A complete solution requires better integration of usage evaluation into other design activities. Without such integration, usability practices will continue to be met often with disappointment, distrust, scepticism and a lack of appreciation in some technology development settings (Iivari 2005).

This sets us up for a third alternative definition of usability that steers a middle course between essentialism and relationalism:

> "Usability is the extent of impact of negative user experiences and negative outcomes on the achievable worth of an interactive system. A usable system does not degrade or destroy achievable worth through excessive or unbearable usage costs."

Usability can thus be understood as a major facet of user experience that can reduce achieved worth through adverse usage costs, but can only add to achieved worth through the iterative removal of usability problems. Usability improvements reduce usage costs, but cannot increase the value of usage experiences or outcomes. In this sense, usability has the same structural position as Herzberg's (Herzberg 1966) *hygiene factors* in relation to his *motivator* factors.

## 15.6.4 From Hygiene Factors to Motivators


FIGURE 15.22: Usability is a negative hygiene factor, not a positive motivator.

*Courtesy of Office for Emergency Management. U.S. Office of War Information. Domestic Operations Branch. Bureau of Special Services. Copyright: pd (Public Domain (information that is common property and contains no original authorship)).*

Herzberg studied motivation at work, and distinguished positive *motivators* from negative *hygiene* factors in the workplace. Overt and sustained recognition at work is an example of a motivator factor, whereas inadequate salary is an example of a hygiene factor. Motivator factors can cause job satisfaction, whereas hygiene factors can cause dissatisfaction. Although referred to as Herzberg's two-factor theory (after the two groups of factors), it spans three valences: positive, neutral and negative. The absence of motivators does not result in dissatisfaction, but in the (neutral) absence of (dis)satisfaction. Similarly, the absence of negative hygiene factors does not result in satisfaction, but in the (neutral) absence of (dis) satisfaction. Loss of a positive motivator thus results in being unsatisfied, whereas loss of an adverse hygiene factor results in being undissatisfied! Usability can thus be thought of as an overarching term for hygiene factors in user experience. Attending to poor usability can remove adverse demotivating hygiene factors, but it cannot introduce positive motivators.

Positive motivators can be thought of as the opposite pole of user experience to poor usability. Poor usability demotivates, but good usability does not motivate, only positive experiences and outcomes do. The problem with usability as originally conceived in isolation from other design concerns is that it only supports the identification and correction of defects, and not the identification and creation of positive qualities. Commercially, poor usability can make a product or service uncompetitive, but usability can only make it competitive relative to products or services with equal value but worse usability. Strategically, increasing value is a better proposition than reducing usage costs in any market where overall usability is 'good enough' across competitor products or services.

## 15.7 FUTURE DIRECTION FOR USABILITY EVALUATION

Usage costs will always influence whether an interactive system is worthwhile or not. These costs will continue to be so high in some usage contexts that the achieved worth of an interactive system is degraded or even destroyed. For the most part, such situations are avoidable, and will only persist when design teams lack critical human-centred competences. While originally encountered in systems developed by software engineers, poor usability is now also linked to design decisions imposed by some visual designers, media producers, marketing 'suits', interfering managers, inept committees, or in-house amateurs. Usability experts will continue to be needed to fix their design disasters.

### 15.7.1 Putting Usability in its Place

In well directed design teams, there will not be enough work for a pure usability specialist. This is evidenced by a trend within the last decade of a broadening from usability to user experience expertise. User experience work focuses on both positive and negative value, both during usage and after it. A sole focus on negative aspects of interactive experiences is becoming rarer. Useful measures of usage are extending beyond the mostly cognitive problem measures of 1980s usability to include positive and negative affect, attitudes and values, e.g., fun, trust, and self-affirmation. The coupling between evaluation and design is being improved by user experience specialists with design competences. We might also include interaction designers with user experience competences, but no interaction designer worthy of the name should lack these! Competences in high-fidelity prototyping, scripting and even programming are allowing user experience specialists firstly to communicate human context insights through working prototypes (Rosenbaum 2008), and secondly to communicate possible design responses to user experience issues revealed in evaluations.

Many user experience professionals have also developed specific competences in areas such as brand experience, trust markers, search experience/optimisation, usable security and privacy, game experience, self and identity, and human values. We can see two trends here. The first involves complementing human-centred expertise with strong understandings of specific technologies such as search and security. The second involves a broadening of human-centred expertise to include business competences (e.g., branding) and humanistic psychological approaches (e.g., phenomenology, meaning and value). At the frontiers of user experience research, the potentials for exploiting insights from the humanities are being increasingly demonstrated (e.g., Bardzell 2011, Bardzell and Bardzell 2011, Blythe et al. 2011).

The extension of narrow usability expertise to broader user experience competences reduces the risk of inappropriate evaluation measures (Cockton 2007). However, each new user experience attribute introduces new measurement challenges, as do longer term measures associated with achieved value and persistent adverse consequences. A preference for psychometrically robust metrics must often be subordinated to the needs to measure specific value in the world, however and wherever it occurs. User experience work will thus increasingly require the development of custom evaluation instruments for experience attributes and worthwhile outcomes. Standard validated measures will continue to add value, but only if they are the right measures. There is however a strong trend towards custom instrumentation of digital technologies, above the level of server logs and low level system events (Rosenbaum 2008). Such custom instrumentation can extend beyond a single technology component to all critical user touch points in its embracing product-service ecosystem. For example, where problems arise with selecting, collecting, using and returning hired vans, it is essential to instrument the van hire depots, not the web site. Where measures relate directly to designed benefits and anticipated adverse interactions, this approach is known as *direct worth instrumentation* (Cockton 2008b).


FIGURE 15.23: The right instrumentation is crucial for worthwhile evaluation.



Risks of inappropriate standard metrics arise when web site evaluations use the ready-to-hand measures of web server logs. What is easy to measure via a web server is rarely what is needed for meaningful relevant user experience evaluation. Thus researchers at Google (Rodden et al. 2010) have been developing a set of more relevant user experience ('HEART') measures to replace or complement existing log-friendly metrics ('PULSE' measures). The HEART measures are Happiness, Engagement, Adoption, Retention, and Task success. The PULSE measures are Page views, Uptime, Latency, Seven-day active users (i.e. count of unique users who used system at least once in last week), and Earnings. All PULSE measures are easy to make, but none are always relevant.

Earnings (sales) can of course be a simple and very effective measure for e-commerce as a measure of not one, but every, user interaction. As an example of the effectiveness of sales metrics, Sunderland University's Alan Woolrych (see Figure 15.7) has contributed his expertise to commercial usability and user experience projects that have increased sales by seven digits (in UK sterling), increasing sales in one case by at least 30%. Improved usability has been only one re-design input here, albeit a vital one. Alan's most successful collaborations involve marketing experts and lead business roles. Similar improvements have been recorded by collaborations involving user experience agencies and consultancies worldwide. However, the relative contributions of usability, positive user experience, business strategy and marketing expertise are not clear, and in some ways irrelevant. The key point is that successful e-commerce sites require *all* such inputs to be co-ordinated throughout projects.

There are no successful digital technologies without what might be regarded as usability flaws. Some appear to have severe flaws, and are yet highly successful for many users. Usability's poor reputation in some quarters could well be due to its focus on the negative at the expense of the positive. What matters is the resulting balance of worth as judged by all relevant stakeholders, i.e., not just users, but also, for example, projects' sponsors, service provision staff, service management, politicians, parents, business partners, and even the general public.


Figure 15.24: Evaluation of the Whereabouts Clock revealed unexpected benefits.


Figure 15.25: The Whereabouts Clock in its usage context.



Evaluation needs to focus on both positives and negatives. The latter need to be identified and assessed for their impact on achieved worth. Where there are unacceptable adverse impacts, re-design and further evaluation is needed to confirm that unintended negative experiences and/or outcomes have been 'designed out'. However, evaluation misses endless opportunities when it fails to identify unintended positives experiences and/or outcomes. Probe studies have proved to be highly effective here, identifying positive appropriative use that was completely unanticipated by design teams (e.g., Brown et al. 2007, Gaver et al. 2008). It is refreshing to encounter evaluation approaches that identify unexpected successes as well as unwanted failures. For example, the evaluation of the Whereabouts

Clock (Brown et al. 2007) revealed one boy's comfort at seeing his separated family symbolically united on the clock's face.

Designers and developers are more likely to view evaluation positively if it is not overwhelmingly negative. Also, this spares evaluators from ritually starting their reports with a 'few really good points about the design' before switching into a main body of negative problems. There should always be genuine significant positive experiences and outcomes to report.

Evaluation becomes more complicated once positive and negative phenomena need to be balanced against each other across multiple stakeholders. *Worth* has been explored as an umbrella concept to cover all interactions between positive and negative phenomena (Cockton 2006). As well as requiring novel custom evaluation measures, this also requires ways to understand the achievement and loss of worth. There have been some promising results here with novel approaches such as *worth maps* (Cockton et al. 2009a, Cockton et al. 2009b, Otero and José 2009). Worth maps can give greater prominence to system attributes while simultaneously relating them to contextual factors of human experiences and outcomes. Evaluation can focus on worth map elements (system attributes, user experience attributes, usage outcomes) or on the connections between them, offering a practical resource for moving beyond tensions between essentialist and relational positions on software quality.

Worth-focused evaluation remains underdeveloped, but will focus predominantly on outcomes unless experiential values dominate design purpose (as in many games). Where experiential values are not to the fore, detailed evaluation of user interactions may not be worthwhile if products and services have been shown to deliver or generously donate value. Evaluation of usage could increasingly become a relatively infrequent diagnostic tool to pinpoint where and why worth is being degraded or destroyed. Such a strategic focus is essential now that we have new data collection instruments such as web logs and eye tracking that gather massive amounts of data. Such new weapons in the evaluation arsenal must be carefully aimed. A 12-bore shotgun scattershot approach cannot be worthwhile for any system of realistic complexity. This is particularly the case when, as in my personal example of phone contacts transfer, whole product ecologies (Forlizzi 2008) must be evaluated, and not component parts in isolation. When usage within such product ecologies here is mobile, intermittent and moves through diverse social contexts, it becomes even more unrealistic to evaluate every second of user interaction.

In the future, usability evaluation will be put in its place. User advocates will not be given free rein to berate and scold. They will become integral parts of design teams with **B**alanced, **I**ntegrated and **G**enerous (BIG!) design practices. It's time for all the stragglers in usability evaluation to catch up with the **BIG** boys and girls. Moaning on the margins about being ignored and undervalued is no longer an option. Usability must find its proper place within interaction design, as an essential part of the team, but rarely King of the Hill. The reward is that usability work could become much more rewarding and less fraught. That has got to be worthwhile for all concerned.

*Courtesy of Joel Rogers. Copyright: pd (Public Domain (information that is common property and contains no original authorship)).*

Figure 15.26 A-B: From Solo Specialist to Team Member: User Experience as an integrated part of design teams.

## 15.8 WHERE TO LEARN MORE

*HCI Remixed* (Erickson and McDonald 2008) is an excellent collection of short essays on classic HCI books and papers, plus other writing that has influenced leading HCI researchers. It contains a short essay (Cockton 2008a) on the Whiteside et al. (1988) classic, and many more of interest.

There is a short account of BIG Design in Cockton, G. Design: BIG and Clever, Interfaces Magazine, 87, British Interaction Group, ISSN 1351-119X 2011, 5-7

Sears' and Jacko's HCI Handbook (Sears and Jacko 2007) is a very comprehensive collection of detailed chapters on key HCI topics. The 3rd edition will be published in 2012. There are chapters on user testing, inspection methods, model-based methods and other usability evaluation topics.

Darryn Lavery prepared a set of tutorial materials on inspection methods in the 1990s that are still available:

- Lavery, D., Cockton, G., and Atkinson, M. P. 1996. Heuristic Evaluation: Usability Evaluation Materials, Technical Report TR-1996-15, University of Glasgow. Accessed 15/9/11 at http://www.dcs.gla.ac.uk/asp/materials/HE_1.0/materials.pdf
- Lavery, D., Cockton, G., and Atkinson, M. P. 1996. Heuristic Evaluation for Software Visualisation: Usability Evaluation Materials, Technical Report TR-1996-16, University of Glasgow, 1996. Accessed 15/9/11 at http://www.dcs.gla.ac.uk/asp/materials/SVHE_1.0/materials.pdf
- Lavery, D., Cockton, G., and Atkinson, M. P. 1996. Cognitive Dimensions: Usability Evaluation Materials, Technical Report TR-1996-17, University of Glasgow. Accessed 15/9/11 at http://www.dcs.gla.ac.uk/asp/materials/CD_1.0/materials.rtf
- Lavery, D., and Cockton, G. 1997. Cognitive Walkthrough: Usability Evaluation Materials, Technical Report TR-1997-20, Department of Computing Science, University of Glasgow. Edited version available 15/9/11 as http://www.dcs.gla.ac.uk/~pdg/teaching/hci3/cwk/cwk.html

Europe's COST programme has funded two large research networks on evaluation and design methods. The MAUSE project (COST Action 294, 2004-2009) focused on maturing usability evaluation methods. The TwinTide project (COST Action

IC0904, 2009-2013) has a broader focus on design and evaluation methods for interactive software. There are several workshop proceedings on the MAUSE web site (www.cost294.org), including the final reports, as well as many publications by network members on the associated MAUSE digital library. The TwinTide web site (www.twintide.org) is adding new resources as this new project progresses.

The Usability Professionals Association, UPA, have developed some excellent resources, especially their open access on-line Journal of Usability Studies. Their Body of Knowledge project, BOK, also has created a collection of resources on evaluation methods that complement the method directory prepared by MAUSE WG1. Practically minded readers may prefer BOK content over more academically oriented research publications.

Jakob Nielsen has developed and championed discount evaluation methods for over two decades. He co-developed Heuristic Evaluation with Rolf Mohlich. Jakob's www.useit.com web site contains many useful resources, but some need updating to reflect some major developments in usability evaluation and interaction design over the last decade. For example, in the final version of his heuristics some known issues with Heuristic Evaluation are not covered. Even so, the critical reader will find many valuable resources on www.useit.com. Hornbæk (2010) is a very good source of critical perspectives on usability engineering, and should ideally be read alongside browsing within www.useit.com.

The American Association for Computing Machinery (ACM) sponsors many key HCI conferences through its SIGCHI special interest group. The annual CHI (Computer-Human Interaction) conference is an excellent source for research papers. There is no specialist ACM conference with a focus on usability evaluation, but the SIGCHI DIS (Designing Interactive Systems) conference proceedings and the DUX (Designing for User Experiences) conference proceedings do contain some valuable research papers, as does the SIGCHI CSCW conference series. The SIGCHI UIST conference (Symposium on User Interface Software and Technology) often includes papers with useful experimental evaluations of innovative interactive components and design parameters. All ACM conference proceedings can be accessed via the ACM Digital Library. Relevant non ACM conferences include UPA (The Usability Professionals' Association international conference), ECCE (the European Conference on Cognitive Ergonomics), Ubicomp (International Conference on Ubiquitous Computing), INTERACT (the International Federation for Information Processing Conference on Human-Computer Interaction) and the British HCI Conference series. UPA has a specific practitioner focus on usability evaluation. Most HCI publications are indexed on www.hcibib.org. In November 2011, a search for *usability evaluation* found almost 1700 publications.

## 15.9 ACKNOWLEDGEMENTS

I have been immensely fortunate to have collaborated with some of the most innovative researchers and practitioners in usability evaluation, despite having no serious interest in usability in my first decade of work in Interaction Design and HCI!

One of my first PhD students at Glasgow University, Darryn Lavery, changed this through his struggle with what I had thought was going to be a straightforward PhD on innovative inspection methods. Darryn exposed a series of serious fundamental problems with initial HCI thinking on usability evaluation. He laid the foundations for over a decade of rigorous critical research through his development of conceptual critiques (Lavery et al. 1997), problem report formats (Lavery and Cockton 1997), and problem extraction methodologies (Cockton and Lavery 1999). From 1998, Alan Woolrych, Darryn Lavery (to 2000), myself and colleagues at Sunderland University built on these foundations in a series of studies that exposed the impact of specific resources on the quality of usability work (e.g., Cockton et al. 2004), as well as demonstrating the effectiveness of these new understandings in Alan's commercial and e-government consultancies. Research tactics from our studies were also used to good effect from 2005-2009 by members of WG2 of COST Action 294 (MAUSE - see *Where to learn more* above), resulting in a new understanding of evaluation methods as usability work that adapts, configures and combines methods (Cockton and Woolrych 2009). COST's support for MAUSE and the the follow on TwinTwide Action (*Where to learn more*, below) has been invaluable for maintaining a strong focus on usability and human-centred design methods in Europe. Within Twintide, Alan Woolrych, Kasper Hornbæk, Erik Frøkjær and I have applied results from MAUSE to the analysis of Inspection Methods (Cockton et al. 2012), and more broadly within the broader context of usability work (Woolrych et al. 2011).

Nigel Bevan, a regular contributor to MAUSE and TwinTide activities has provided helpful advice, especially on international standards. Nigel is one of many distinguished practitioners who have generously shared their expertise and given me feedback on my research. At the risk of omission and in no particular order, I would particularly like to acknowledge generous sharing of knowledge of and insights on usability and emerging evaluation practices by Tom Hewett, Fred Hansen, Jonathan Earthy, Robin Jeffries, Jakob Nielsen, Terry Roberts, Bronwyn Taylor, Ian McClelland, Ken Dye, David Caulton, Wai On Lee, Mary Czerwinski, Dennis Wixon, Arnie Lund, Gaynor Williams, Lynne Coventry, Jared Spool, Carolyn Snyder, Will Schroeder, John Rieman, Giles Colborne, David Roberts, Paul Englefield, Amanda Prail, Rolf Mohlich, Elizabeth Dykstra-Erickson, Catriona Campbell, Manfred Tscheligi, Verena Giller, Regina Bernhaupt, Lucas Noldus, Bonnie John, Susan Dray, William Hudson, Stephanie Rosenbaum, Bill Buxton, Marc Hassenzahl, Carol Barnum, William Hudson, Bill Gaver, Abigail Sellen, Jofish Kaye, Tobias Uldall-Espersen, John Bowers and Elizabeth Buie. My apologies to anyone who I have left out!

## 15.10 COMMENTARY BY DAVID A. SIEGEL

How to cite this commentary in your report

### David A. Siegel

*© David A. Siegel*

David is a well-known user experience researcher and consultant, and co-owner of Dray & Associates, Inc. He specializes in using field user studies and contextual research, and naturalistic usability evaluation to help guide product concept, strategy, and interaction design. He has published and taught on a variety of user-centered design topics, including many workshops and tutorials ...

David A. Siegel
David A. Siegel is a member of The Interaction Design Foundation

I appreciate the opportunity to comment on Gilbert Cockton's chapter on usability. My comments come from the perspective of someone who has practiced user experience (UX) research of many types as a consultant. Although I have done my share of usability evaluations, almost all of my work currently consists of in vivo contextual research, with a focus on finding ways to increase value to the user. The product teams I work with often include internal usability specialists, and I well understand their roles within their teams and the challenges they face. Finally, my prior career as a psychologist has given me a very healthy respect for the difficulties of measuring and understanding human behavior in a meaningful way, and impatience with people who gloss over these challenges.

To begin with points of agreement, I applaud Gilbert's emphasis on the need to consider usability in the context of all other factors that influence the value people obtain from interactive products. I also agree with his critique of the methodological limitations of laboratory usability evaluation. I could not agree more that contextual research is usually much more powerful than laboratory usability evaluation as an approach to understanding the user experience holistically and to gaining insights that will drive UX design towards greater overall value. I also agree with Gilbert's call to usability professionals to focus on the larger issues.

With this said, however, I have a number of concerns about the chapter's portrayal and critique of usability as an inherently limited, marginal contributor to development of great products. In regard to practice, there are many gradations of skill and wisdom, and some unknown proportion of usability practitioners may deserve to be confronted with the criticisms Gilbert raises. However, I question the idea that these criticisms are true of usability practice in principle. I believe that most mature usability practitioners are aware of the issues he raises, would agree with many of his points, and work hard to address them in various ways. In the discussion that follows, I will present an alternate view of usability's role as a fundamental contributor to product value. This requires considering usability at two levels: as an abstract concept and as a field of practice.

First, one note on terminology: throughout this commentary I use the word "product" to refer to anything that is being designed for interactive use, be it software, website, system, or device, or any new features of these.

### 15.10.1 Usability and Value as Abstract Constructs

It has become commonplace to emphasize a distinction between usability and value, and also to claim that "experience" has superseded usability. This treats usability as though it is distinct from both of these other concepts. Even though usability is generally acknowledged to be important, it is portrayed as quite subordinate. In Gilbert's chapter, this is reflected in the idea that usability is merely a "hygiene factor," the absence of which can block the delivery of value or reduce it by adding to costs, but one which can never go beyond neutral as a contributor to value. In my view, this greatly understates the contribution of usability to value. The two concepts are far more intertwined than this Attempts to abstract value from usability are just as flawed as the reverse.

The notion that ease of use is a separate issue from value, although one that affects it, has much face validity. It seems to make sense to think of value as a function of benefit somehow related inversely with costs, with usability problems counted in the costs column. Unfortunately, this is consistent with the notion of usability as "a feature," something that makes usability professionals cringe, just as the idea of design as the "lipstick" applied to a product in the last stage makes designers cringe. In my view, usability divorced from value is as undefined as the sound of one hand clapping. Usability can only be defined in the context of benefit. By this I do not mean benefit in principle, but rather the benefit anticipated by or experienced by the user. At one level, this is because usability and experienced benefit interact in complex ways. But beyond this, there are many products where usability is itself the primary value proposition. In fact, the central value proposition of most technological tools is that they make something of value easier to achieve than it used to be. A mobile phone has value because its portability enables communication while mobile, and its portability matters because it makes it more usable when mobile.

In another example, a large medical organization I am familiar with recently adopted a new, integrated digital medical record system. Initially, there was a great deal of grumbling about how complex and confusing it was. I saw the classic evidence of problems in the form of notes stuck on computer monitors warning people not to do seemingly intuitive things and reminding them of the convoluted workarounds. However, more recently, I have heard nurses make comments about the benefit of the system. Doctors' orders are entered electronically and made automatically available to the appropriate departments. As a result, patients now can come to the clinic for a follow up laboratory test without having to remember to bring a written copy of the lab order. "Usability" is not simply the issue of whether doctors can figure out how to enter the order in the system and direct it to the lab rather than the ophthalmology department, although that is part of it. The benefit has to do with its overall success in reducing the usability problems of an earlier process that used to be difficult to coordinate and error prone, and this increase in usability only matters because it is delivering a real benefit.


DELL
When opening Anesthesia
Manager - only click
ONCE! (+ be patient)
Look for AM at bottom of
screen
If you click AM more than
once, EPIC will not open!

FIGURE 15.1

Sometimes, usability seems detached from value when the goal is fulfilled at the end of a sequence of steps, but the steps along the way are confusing. However, it can be the separation from the experience of value that creates the usability problem. For example, if people trying to book an online hotel reservation get lost in preliminary screens where they first have to create an account, we might see usability as only relevant to the cognitive aspects of the sign up process, and as mere hygiene factors. But when users become disoriented because they do not understand what a preliminary process has to do with their goal, it can be precisely because they cannot see the value of the preliminary steps. That is, they can't see how the steps contribute to something they care about, and lead them towards their goal. If they did, the subparts of the process would both be more understandable and would acquire value of their own, just as a well-designed hammer gains value not simply in its own right, but because it is understood as a more effective tool for driving nails (which are valued because of the value of the carpentry tasks they enable, and so on.) This is simultaneously a usability problem and an "experience of value problem." For this reason, a common challenge of usability is to convey to users that they are making progress towards an outcome that they value.

For example, in one product that I worked on, users were offered the opportunity to enroll for health insurance benefits that claimed to be highly personalized. In addition to setting different benefit levels for different members of their families, users could compose their own preferred networks of medical specialists, for which they would receive the highest reimbursement levels. Unfortunately, the actual user experience did not appear to live up to this. As soon as the user entered identifying information, the system applied defaults to all the decisions that the user was supposedly able to personalize. It only fulfilled its value proposition of personalization by allowing the user to "edit" the final configuration—13 screens into the process. Along the way, the user experienced the sense that decisions were being imposed. There was not even an indication to the user that the opportunity to make personal choices was coming eventually. Unfortunately, the system did not start by asking the user which choices mattered to them and what their preferences were, so it could factor this things in before presenting a result to the user.

How should we construe this? As a usability problem? As a problem in delivery of value? As a failure in the design of a user experience? It is all of these at the same time. The discrepancy from the expected perception of value is a primary cause of the confusion users felt. None of these constructs (usability, value, experience) can be defined without incorporating the others. If we parse and remove the meaning that we can attribute to any of them, we drain the meaning from the others. Disputes about which is the legitimate language to describe them are at best just ways to emphasize different faces of the same phenomenon, and at worst semantic quibbling. This means that usability is something more than just another item to add into the costs column when we weigh them against benefits to arrive at value. It also means we can't answer the question of whether something is usable without also answering the question, "What matters?"

### 15.10.2 Usability Practice in Product Development

While Gilbert and I may agree on the need for a more holistic focus on user experience, we may disagree about whether usability in practice actually takes this holistic view. Reducing the profession to a particular type of laboratory evaluation makes it seem limited and can raise questions about its relevance. While as I said, I agree with Gilbert's critique of the methodological limitations of this approach, the profession is far broader and more diverse than this. Furthermore, even despite its limitations, traditional usability evaluation often contributes significant value in the product development context, at least when practiced by reflective professionals. Below, I comment on some of the major issues Gilbert raises with regard to usability practice.

#### 15.10.2.1 Is 'Ease of Use' still relevant?

Although some interaction design patterns have become established, and an increasing number of users have gained generalizable skills in learning a variety of new interaction patterns, this does not mean that ease of use as an issue has gone away or even declined in importance. For several reasons, it makes more sense to see the spectrum of usability issues to be addressed as having evolved. First, the spectrum of users remains very large and is constantly expanding, and there are always some at an entry level. Second, although with experience users may gain knowledge that is transferrable from one family of products to another, this can be both an asset and a source of confusion, because the analogies among product designs are never perfect. Third, as innovation continues to create new products with new capabilities, the leading edge of UX keeps moving forward. On that leading edge, there are always new sets of design challenges, approaches, and tradeoffs to consider. Finally, the world does not consist only of products intended to create experiences for their own sake as opposed to those that support tasks (a distinction that is not necessarily so clear). Products that are designed to facilitate and manage goal-oriented tasks and to support productivity continue to have a tremendous impact on human life, and we have certainly not learned to optimize ease of interaction with them. Finally, usability is continually driven forward by competition within a product domain.

Another claim in the chapter that suggests limited relevance for usability is that good product teams do not need a dedicated usability person. This is too simplistic. Of course, a designated usability person does not create usability single handedly. That is the cumulative result of everything that goes into the product. However, how much specialized work there is for a usability person depends on many factors. We need to take into account the variability among ways that product teams can be structured, the magnitude of the UX design challenges they face in their product space, the complexity of the product or family of inter-related products that the usability person supports, how incremental versus innovative the products are, what the risk tolerance is for usability problems, how heterogeneous the user population and user contexts are, how much user persistence is needed for usage to be reinforced and sustained by experiences of value, etc. The simplistic statement certainly does not address the fact that some usability work takes more effort to carry out than others. To do realistic research with consumers is generally much easier than doing realistic research inside enterprises.

As a matter of fact, in actual practice teams often do not have usability professionals assigned to them full time, because these people often support multiple product teams, in a matrix organizational structure. There are benefits to this in terms of distributing a limited resource around the company. But there are also drawbacks. This structure often contributes to the usability person being inundated with requests to evaluate superficial aspects of design. It can also exclude the usability person from integrative discussions that lead to fundamental aspects of product definition and design and determine the core intendedvalue of the product. Some usability people may accept this limited role complacently and passively respond to team requests, in the hopes of providing "good service," but many others recognize the challenges of this role structure and work very hard to get involved with deeper issues of value, exactly as Gilbert urges them to.

### 15.10.2.2 Do usability professionals only focus on cognition?

Several points in Gilbert's critique of practice are based on a limited view of what usability people do. It is true that laboratory usability evaluation typically does try to isolate cognitive factors by treating the users goals and motivation as givens, rather than attempting to discover them. Often, it is the fit of the assumed goal that is in question, and that makes the biggest difference in user experience.

But many usability professionals spend a great deal of time doing things other than laboratory tests, including, increasingly, fundamental in context user research. For many years, usability evaluation has served as a platform to promote systematic attention to deeper issues of value to the user. Many usability professionals deeply understand the complex, entangled relationship between ease of use and value, and work to focus on broad questions of how technology can deliver experienced value. Some usability people have succeeded in getting involved earlier in the design process when they can contribute to deeper levels of decision-making. This has led to their involvement in answering questions about value, like "What will matter to the user?" or "What will influence whether people will really adopt it?" rather than only asking, "Could the user, in principle, figure out how to do it if they wanted to?" There are certainly people who are narrow specialists in a particular set of techniques focused on ease of use, but they do not own the definition of the field, and specialization per se is not bad.

#### 15.10.2.3 What can usability people contribute?

Gilbert is correct that UX skills are increasingly distributed across roles. He lists a number of such skills, but missing from the list is the skill of doing disciplined research to evaluate evidence for the assumptions, claims, beliefs, or proposed designs of the product team, whether these are claims about what people need and will value, or whether a particular interface design will enable efficient performance.

Gilbert points out that there is no cookbook of infallible usability approaches. This is not a surprise, and indeed, we should never have expected such a thing. Such cookbooks do not exist for any complex field, and there is no way to guarantee that a practical measurement approach captures the core meaning of a complex construct. I do agree wholeheartedly with Gilbert when he points out the many factors that can complicate the process of interpreting usability findings due to this lack of a cookbook of infallible methods and the presence of many confounds. These issues argue for the need for greater professionalism among usability practitioners, not for the downgrading of the profession or marginalizing it on the periphery of the product development team. Professionalism requires that practitioners have expert understanding of the limitations of methods, expertise in modifying them to address different challenges, the dedication to continually advance their own processes, and the skill to help drive the evolution of practice over time. At a basic level, mature usability professionals recognize that results from a single evaluation do not give an absolute measure of overall usability. They are careful about overgeneralizing. They at least attempt to construct tasks that they expect users will care about, and attempt to recruit users who feel will engage realistically with the tasks. They wrestle with how best to achieve these things given the constraints they work under. Those who do not recognize the challenges of validity, or who apply techniques uncritically are certainly open to criticism, or should be considered mere technicians, but, again, they do not represent the best of usability practice.

In the absence of scientific certainty, where is the value of usability practice? In the product development context, this should not be judged by how well usability meets criteria of scientific rigor. It is more relevant to ask how it compares to and compliments other types of evidence that are used as a basis for product definition, audience targeting, functional specification, and design decisions. This means we need to consider usability's role within the social and political processes of product development.

Membership in product teams often requires allegiance to the product concept and design approach. Sometimes, demonstrations of enthusiasm are a prerequisite for hiring. Often, it is risky for team members to challenge the particular compromises that have been made previously to adapt the product to various constraints or a design direction that has become established, since these all have vested interests behind them. In this context, the fact that usability methods (or approaches as Gilbert rightfully calls them) are scientifically flawed does not mean they are without value. It is not as though all the other streams of influence that affect product development are based on solid science while usability is voodoo. When you consider the forces that drive product development, it is clear that subjective factors dominate many of them, for example:

- Follow the leader design mentality
- Imperfect and sometimes cherry-picked market research data
- Internal politics
- Impressions derived from other faulty approaches to testing the product's appeal

- Beliefs about who future users will be

Product decisions are also deeply influenced by legitimate considerations that are difficult to evaluate objectively, much less to weigh against each other, such as:

- How to deal with legacy issues and the difficulty of innovating with an installed base
- How to weigh engineering feasibility and manufacturing cost against other considerations
- Market timing
- The business need to differentiate from the competition, even if the competition has a better approach
- The need to promote the inherent advantages of your technical strengths and downplay its limitations

In this context, a discipline that offers structured and transparent processes for introducing evidence-based critical thinking into the mix adds value, even though its methods are imperfect and its evidence open to interpretation. Sometimes, usability evaluation is a persuasive tool to get product teams to prioritize addressing serious problems that everyone knew existed, but that could not receive focus earlier. Sometimes this is needed to counterbalance the persuasive techniques of other disciplines, which may have less scientific basis than usability. Sometimes usability results provide a basis to resolve disputes that have no perfect answer and that have previously paralyzed teams. And sometimes they have the effect of triggering discussions about controversial things that would otherwise have been suppressed.

#### 15.10.2.4 Does usability contribute to innovation?

Sometimes, usability in practice is portrayed as a mere quality assurance process, or as Gilbert says, a hygiene factor. It is often equated with evaluation as distinct from discovery and idea generation. In many ways, this is a false distinction.

Careful evaluation of what exists now can inspire invention and direct creativity towards things that will make the most difference. Practices like rapid iterative design reflect efforts to integrate evaluation and invention. Practices that are considered to be both discovery and invention processes, like contextual design, fall on a continuum with formative usability evaluation and naturalistic evaluation in the usage context. Of course, usability professionals differ in their skills for imagining new ways of meeting human needs, envisioning new forms of interactive experience, or even generating multiple alternative solutions to an information architecture problem or interface design problem. Some may lack these skills. However, the practice of usability is clearly enhanced by them. Those who can integrate evaluation and invention can add more value to the product development process and can help ensure usability/value in the ultimate product.

### 15.10.3 Conclusion

Certainly one can find examples of bad usability practice, and I cannot judge what other people may have encountered. Of course, there is also a lot of bad market research, bad design, bad business decision-making, bad engineering, and bad manufacturing. Let us not define the field based on its worst practice, or even on its lowest-common denominator practice. Failure to take into account the kinds of confounds Gilbert identifies is indeed bad practice because it will lead to misleading information. Handing over to a team narrow findings, minimally processed, excludes the usability practitioner from the integrative dialogue in which various inputs and courses of action are weighed against each other, and from the creative endeavor of proposing solutions. This will indeed limit usability practitioners to a tactical contributor role and will also result in products that are less likely to provide value for the users.

Finally, to any usability practitioners who think that usability is some kind of essence that resides in a product or design, and that can be objectively and accurately measured in the lab: Stop it. If you think that there is a simple definition of ease of use that can be assessed in an error-free way via a snapshot with an imperfect sample of representative users and simulated tasks: Stop it. If you think usability does not evolve over time or interact with user motivation and expectations and experience of benefit: Stop it. If you think that ease of use abstracted from everything else is the sole criterion for product success or experienced value: Stop it! If you think you are entitled to unilaterally impose your recommendations on team decision-making: Stop it. You are embarrassing the profession!

## 15.11 COMMENTARY BY STEPHANIE ROSENBAUM

How to cite this commentary in your report

### Stephanie Rosenbaum

*© Stephanie Rosenbaum*

Stephanie Rosenbaum is founder and president of TecEd, Inc., a 15-person consultancy specializing in usability research and information design. Now in its 41st year of business, TecEd maintains four offices: Ann Arbor, Michigan; Palo Alto, California; Rochester, New York; and Milwaukee, Wisconsin. Clients include Microsoft, Cisco Systems, Sun Microsystems, Yahoo!, eBay, ...

Stephanie Rosenbaum
Stephanie Rosenbaum is a member of The Interaction Design Foundation

Gilbert Cockton's chapter on Usability Evaluation includes a great deal of valuable, interesting, and well-reasoned information. However, its presentation and focus could be more helpful to the Interaction-Design.org audience. If some of this audience consists of practitioners—especially less-experienced practitioners—then Cockton is not speaking to their needs.

Who are you, the readers of this chapter? One of Interaction-Design.org's tag lines says "making research accessible," and its mission statement talks about producing top-grade learning materials to benefit industry and academia. It seems likely that many of you are practitioners in business, technology, healthcare, finance, government, and other applied fields.

As founder and CEO of a user experience consultancy, I find that most people—in both industry and academia—want to learn about usability evaluation as part of their goal to design better products, websites, applications, and services. Especially in industry, philosophical debates about points of definition take second place to the need to compete in the marketplace with usable, useful, and appealing products.

This is not a new observation. As early as 1993, Dumas and Redish [1] pointed out that we don't do usability testing as a theoretical exercise; we do it to improve products. Unfortunately, Cockton loses sight of this key objective and instead forces his readers to follow him as he presents, and then demolishes, an increasingly complex series of hypotheses about the meaning of usability. The danger of this approach is that a casual reader—especially one with a limited command of English—may learn from the chapter precisely the ideas Cockton eventually disproves.

For example, Cockton begins his chapter with several "ideal" propositions about usability as an inherent property of software that can be measured accurately by well-defined methods, regardless of the context of use. Yet as he states later in the chapter, the contextual nature of design—and thus usability—has long been known, not only in the 1988 Whiteside at al. publication Cockton mentions, but also in the work of Gould and Lewis in the 1970s, published in their seminal 1985 article [2].

Throughout his chapter, Cockton continues to build and revise his definitions of usability. The evolution of these definitions is interesting to me personally because of my academic degrees in the philosophy of language. But reading this chapter gives my colleagues in industry only limited help in their role as user experience practitioners conducting usability evaluations of products under development.

In Section 15.1.1—and implicitly throughout the chapter—Cockton associates usability primarily with interactive software. The concept of usability has never applied only to software; ease of use is important to all aspects of our daily life. In 1988, Don Norman wrote about the affordances of door handles [3]. Giving a guest lecture on usability evaluation, I was surprised and impressed by an attendee's comment describing how his company conducted usability testing of electric table saws.

In Section 15.1.2, Cockton describes "a dilemma at the heart of the concept of usability: is it a property of systems or a property of usage?" Why can't it be both? Interactive systems are meaningless without users, and usage must be of something.

The discussion of damaged merchandise (invalid usability methods) in Section 15.2.1 misses the point that most usability work involves applied empirical methods rather than formal experiments. There are—and will always be—evaluator effects in any method which has not been described in enough detail to replicate it. The fact that evaluator effects exist underlines the importance of training skilled evaluators.

I am concerned that Cockton is emphasizing a false dichotomy when he says, "If software can be inherently usable, then usability can be evaluated solely through direct inspection. If usability can only be established by considering usage, then indirect inspection methods (walkthroughs) or empirical user testing methods must be used to evaluate."

There need not be a dichotomy between essentialist ontologies and relational ontologies of usability as described in Section 15.2.4—and it's not clear that this classification adds to the reader's understanding of usability evaluation. Rather, if enough people in enough different contexts have similar user experiences, then guidelines about how to improve those experiences can be created and applied effectively, without using empirical methods for every evaluation.

Also, from a practical standpoint, it is simply not realistic to usability test every element of every product in all of its contexts. A sensible model is to include both inspection/heuristics and empirical research in a product development program, and move back and forth among the methods in a star pattern similar to the star life cycle of interactive system development, with its alternating waves of creative and structuring activities [4]. See Figures 1 and 2.

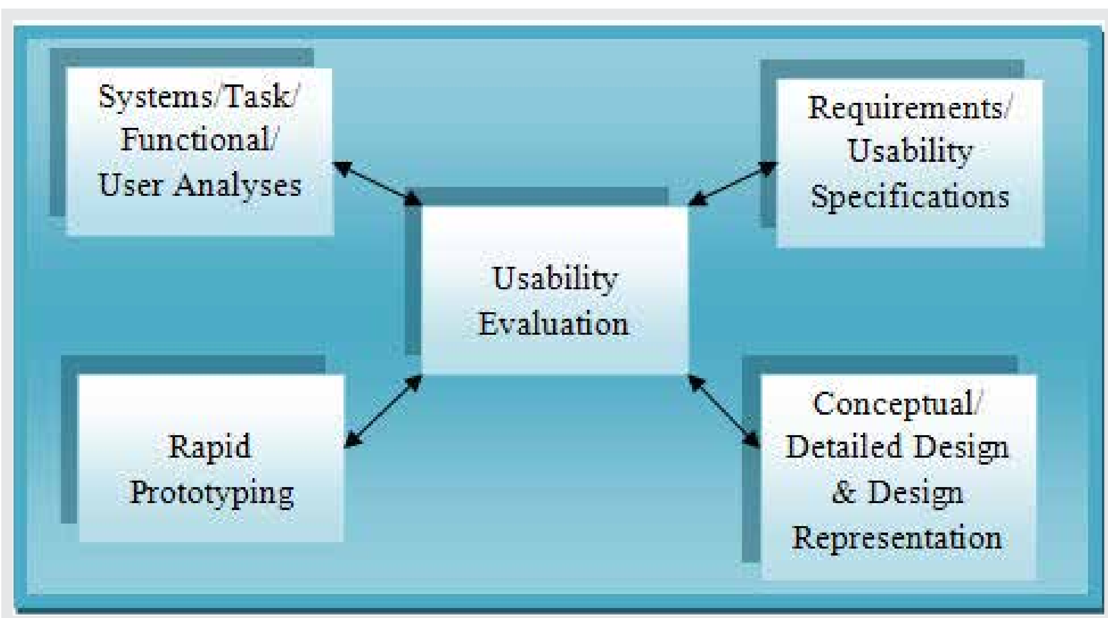



FIGURE 15.1: Star Life Cycle of Interactive System Development (adapted from Hix and Hartson, 1993).

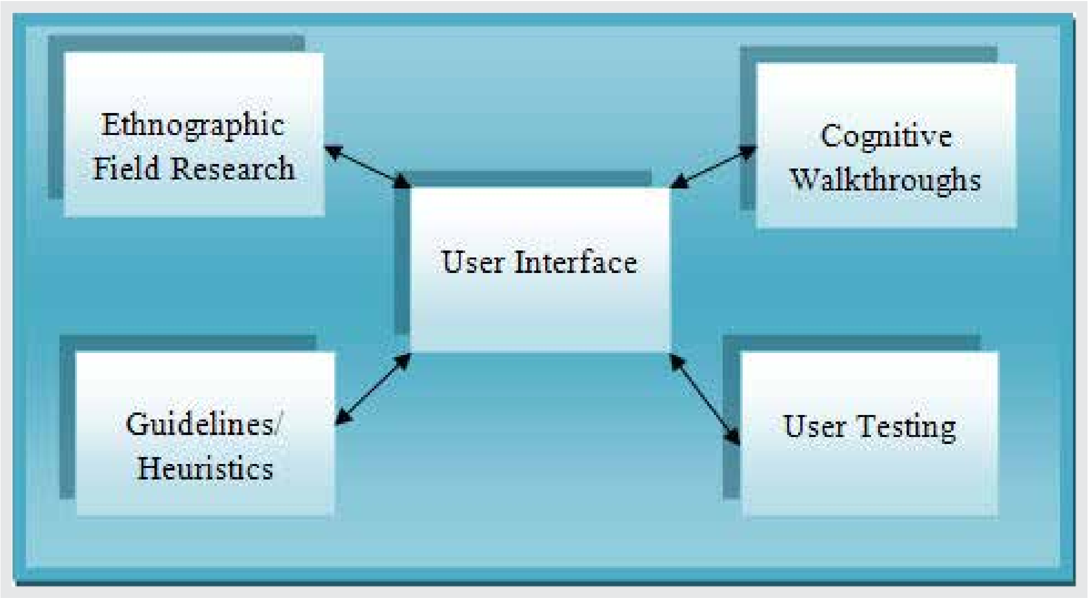



Figure 15.2: Usability Evaluation Star Model.



Thus a key element of usability evaluation is deciding when to employ guidelines and inspection (user-free methods) and when it's critical to perform empirical research such as usability testing or contextual inquiry with the target audience. Planning the activities in a usability evaluation program—and the schedule and budget appropriate to each—is central to the responsibilities of an experienced and skilled usability practitioner. An encyclopedia chapter on usability evaluation should help readers understand this decision-making process.

By the time we get to sections 15.5.1 and 15.5.2, Cockton is accurately describing the situation usability practitioners face: "There is no complete published user testing method that novices can pick up and use 'as is'. All user testing requires extensive project-specific planning and implementation. Instead, much usability work is about configuring and combining methods for project-specific use."

It's true that practitioners in industry, who perform most of today's usability work, typically do not have time or resources to describe their methods in as much detail as do academic researchers. I wish this chapter had provided more references about how to learn usability evaluation skills; adding such a focus would make it more valuable for readers. (I have included a selection of these at the end of my commentary.) Although Cockton correctly points out that such resources are not sufficiently complete to follow slavishly, they are still helpful learning tools.

From my own experience at TecEd, the selection and combination of methods in a usability initiative are the most challenging—and interesting—parts of our consulting practice. For example, our engagements have includeDd the following sequences:

| | |
|---|---|
| **Comcast** | Qualitative research over a four-week period to learn about customers' enjoyment of Comcast Video Instant Messaging and the ease of its use over time, as well as feature and guest-service preferences. The longitudinal study included three phases—we observed and interviewed pairs of Comcast customers, first in their own homes during in-home installation, then in the usability lab to collect more structured behavioral data in a controlled environment, and finally in focus groups to collect preference data after a month's experience with the new service. |
| **Ford Motor Company** | Ethnographic interviews at the homes of 19 vehicle owners throughout the United States. We observed vehicle records and photographed and analyzed artifacts (see Figure 3) to learn how Web technology could support the information needs of vehicle owners. Next we conducted interviews at the homes of 10 vehicle buyers to learn what information they need to make a purchase decision, where they find it, and what they do with it. We subsequently conducted another cycle of interviews at the homes of 13 truck buyers to learn similar information, as well as how truck buyers compare to other vehicle buyers. |

| | |
|---|---|
| **Philips Medical Systems** | Multi-phase qualitative research project with physicians and allied health personnel during the alpha test of a clinical information system at a major U.S. hospital. After initial "out of box" usability testing at the hospital, we coordinated audiotape diary recording and conducted weekly ethnographic interviews, then concluded the project with a second field usability test after six weeks. |
| **A Major Consumer Electronics Company** | Unmoderated card sorting, followed by an information architecture (IA) exploration to help define the user interface for a new product. We began with a two-hour workshop to brainstorm terms for the card sorting, then created and iterated lists of terms, and launched the sorting exercise. For the qualitative IA exploration, we emulated field research in the usability laboratory, a methodology for gaining some benefits of ethnography when it isn't practical to visit users in the field. We used stage design techniques to create three "environments": home, office, and restaurant (see Figure 4). In these environments, we learned some contextual information despite the lab setting. |

| | |
|---|---|
| **Cisco Systems** | Early field research for the Cisco Unified Communications System, observing how people use a variety of communication methods and tools in large enterprise environments. We began each site visit with a focus group, then conducted contextual inquiry with other participants in their own work settings. Two teams of two researchers (one from TecEd, one from Cisco) met in parallel with participants, to complete each site visit in a day. After all the visits, Cisco conducted a full-day data compilation workshop with the research teams and stakeholders. Then TecEd prepared a project report (see Figure 5) with an executive summary that all participating companies received, which was their incentive to join the study. |

## 15.12 COMMENTARY BY ANN BLANDFORD

How to cite this commentary in your report

### Ann Blandford

© *Ann Blandford*

Ann Blandford is Professor of Human–Computer Interaction in the Department of Computer Science at University College London, and served as Director of UCL Interaction Centre (UCLIC) (2004-2011). Her teaching includes User-Centred Evaluation Methods on the MSc in HCI with Ergonomics at UCL. She started her career in industry, as a software engineer, but soon moved into academia, where she...

Ann Blandford
Ann Blandford is a member of The Interaction Design Foundation

Gilbert Cockton's article on Usability Evaluation does a particularly good job of drawing out the history of "usability" and "user experience" (UX), and highlighting the limitations as well as the importance of a classical "usability" perspective. For several years, I taught a course called "Usability Evaluation Methods", but I changed the name to "User-centred Evaluation Methods" because "usability" had somehow come to mean "the absence of bad" rather than "the presence of good". Cockton argues that "user experience" is the more positive term, and we should clearly be aiming to deliver systems that have greater value than being "not bad".

However, there remains an implicit assumption that evaluation is summative rather than formative. For example, he discusses the HEART measures of Happiness, Engagement, Adoption, Retention and Task success, and contrasts these with the PULSE measures. Used effectively, these can give a measure of the quality (or even the worth) of a system, alone or in the product ecologies of which it is a part. However, they do not provide information for design improvement. A concern with the quantifiable, and with properties of evaluation methods such as reliability (e.g. Hertzum & Jacobsen, 2001), has limited our perspective in terms of what is valuable about evaluation methods. Wixon (2003) argues that the most important feature of any method is its downstream utility: does the evaluation method yield insights that will improve the design? To deliver downstream utility, the method has to deliver insights not just about *whether* a product improves (for example) user happiness, but also *why* it improves happiness, and *how* the design could be changed to improve happiness even further (or reduce frustration, or whatever). This demands evaluation methods that can inform the design of next-generation products.

Of course, no method stands alone: a method is simply a tool to be used by practitioners for a purpose. As Cockton notes, methods in practice are adopted and adapted by their users, so there is in a sense no such thing as a "method", but a repertoire of resources that can be selected, adapted and applied, with more or less skill and insight, to yield findings that are more or less useful. To focus this selection and adaptation process, we have developed the Pret A Rapporter framework (Blandford et al, 2008a) for planning a study. The first important element of the framework is making explicit the obvious point that every study is conducted for a purpose, and that that purpose needs to be clear (whether it is formative or summative, focused or exploratory). The second important element is that every study has to work with the available resources and constraints: every evaluation study is an exercise in the art of the possible.

Every evaluation approach has a potential scope — purposes for which it is and is not well suited. For example, an interview study is not going to yield reliable findings about the details of people's interactions with an interface (simply because people cannot generally recall such details), but might be a great way to find out people's attitudes to a new technology; a GOMS study (John and Kieras, 1996) can reveal important points about task structure, and deliver detailed timing predictions for well structured tasks, but is not going to reveal much about user attitudes to a system; and a transaction log analysis will reveal what people did, but not why they did it.

Cockton draws a distinction between analytical and empirical methods, where analytical methods involve inspection of a system and empirical methods are based on usage. This is a good first approximation, but hides some important differences between methods. Some analytical methods (such as Heuristic Evaluation or Expert Walkthrough) have no direct grounding in theory, but provide more or less support for the analyst (e.g. in the form of heuristics); others (including GOMS) have a particular theoretical basis which typically both constrains the analyst, in terms of what issues can be identified through the method, and provides more support, yielding greater insight into the underlying causes of any issues identified, and hence a stronger basis to inform redesign. In a study of several different analytical methods (Blandford et al, 2008c), we found that methods with a clear theoretical underpinning yielded rich insights about a narrow range of issues (concerning system design, likely user misconceptions, how well the system fits the way users think about their activities, the quality of physical fit between user and system, or how well the system fits its context of use); methods such as Heuristic Evaluation, which do not have theoretical underpinnings, tend to yield insights across a broader range of issues, but also tend to focus more on the negative (what is wrong with a system) than the positive (what already works well, or how a system might be improved).

Cockton rightly emphasises the importance of context for assessing usability (or user experience); surprisingly little attention has been paid to developing methods that really assess how systems fit their users in their various contexts of use. In the context of e-commerce, such as his van hire example, it is widely recognised that the Total Customer Experience matters more than the UX of the website interface (e.g. Minocha *et al,* 2005): the website is one component of a broader system, and what matters is that the whole system works well for the customers (and also for the staff who have to work within it). The same is true in most contexts: the system has to perform well, it has to be usable and provide a positive user experience, but it also has to fit well into the context of use.

In different contexts, different criteria become prominent. For example, for a banking system, security is at least as important as usability, and having confidence in the security of the system is an important aspect of user experience. A few days ago, I was trying to set up a new standing order (i.e. regular payment from my bank account to a named payee) to pay annually at the beginning of the year ... but the online banking system would only allow me to set up a new standing order to make a payment in the next four months, even though it would permit payment to be annual. This was irritating, and a waste of time (as I tried to work out whether there was a way to force the system to accept a later date for first payment), but it did not undermine my confidence in the system, so I will continue to use it because in many other situations it provides a level of convenience that old-fashioned banking did not.

Cockton points out that there are many values that a system may offer other than usability. We have recently been conducting a study of home haemodialysis. We had expected basic usability to feature significantly in the study, but it does not: not because the systems are easy to use (they are not), but because the users have to be very well trained before they are able to dialyse at home, their lives depend on dialysis (so they are grateful to have access to such machines), and being able to dialyse at home improves their quality of life compared to having to travel to a dialysis centre several times a week. The value to users of usability is much lower than the values of quality of life and safety.

Particularly when evaluating use in context, there doesn't have to be an either-or between analytical and empirical methods. In our experience, combining empirical studies (involving interviews and observations) with some form of theory-based analysis provides a way of generalising findings beyond the particular context that is being studied, while also grounding the evaluation in user data. If you do a situated study of (for example) a digital library in a hospital setting (Adams et al, 2005), it is difficult to assess how, or whether, the findings generalise to even a different hospital setting, never mind other contexts of use. Being able to apply a relevant theoretical lens (in this case, Communities of Practice) to the data gives at least some idea of what generalises and what doesn't. In this case, the theory did not contribute to an understanding of usability *per se*, but to an understanding of how the deployment of the technology influenced its acceptance and take-up in practice. Similarly, in a study of an ambulance dispatch system (Blandford and Wong, 2004), a theory of situation awareness enabled us to reason about which aspects of the system design, and the way it was used in context, supported or hindered the situation awareness of control room staff. It was possible to apply an alternative theoretical perspective (Distributed Cognition) to the same context of use (ambulance dispatch) (Furniss and Blandford, 2006) to get a better understanding of how the technology design and workspace design contribute to the work of control room staff, including the ways that they coordinate their activity. By providing a semi-structured method (DiCoT) for conducting Distributed Cognition analyses of systems (Blandford and Furniss, 2006), we are encoding key aspects of the theory to make it easier for others to apply it (e.g. McKnight and

Doherty, 2008), and we are also applying it ourselves to new contexts, such as an intensive care unit (Rajkomar and Blandford, in press). Even though particular devices are typically at the centre of these studies, they do not focus on classical usability of the device, or even on user experience as defined by Cockton, but on how the design of the device supports work in its context of use.

Another important aspect of use in context is how people think about their activities and how a device requires them to think about those activities. Green (1989) and others (Green et al, 2006) developed Cognitive Dimensions as a vocabulary for talking about the mismatch between the way that people conceptualise an activity and the way they can achieve their goals with a particular device; for example, Green proposes the term "viscosity" to capture the idea that something that is conceptually simple (e.g. inserting a new figure in a document) is practically difficult (requiring each subsequent figure to be renumbered systematically in many word processors). We went on to develop CASSM (Blandford et al, 2008b) as a method for systematically evaluating the quality of the conceptual fit between a system and its users. Where there are different classes of users of the same system, which you might regard as different personas, you are likely to find different qualities of fit (Blandford et al, 2002). CASSM contrasts with most established evaluation methods in being formative rather than summative; in focusing on concepts rather than procedures; in being a hybrid empirical-analytical approach; and in focusing on use in context rather than either usability or user experience as Cockton describes them. It is a method for evaluating how existing systems support their users in context, which is a basis for identifying future design opportunities to either improve those systems or deliver novel systems that address currently unmet needs. Evaluation should not be the end of the story: as Carroll and Rosson (1992) argue, systems and uses evolve over time, and evaluation of the current generation of products can be a basis for designing the next generation.

This commentary has strayed some way from the classical definitions of usability as encapsulated in many of the standards, and cited by Cockton, to focus more on how to evaluate "quality in use", or the "extent to which a product can be used by specified users to achieve specified goals" *within their situated context of use*. Cockton argues that "several evaluation and other methods may be needed to identify and relate a nexus of causes". I would argue that CASSM and DiCoT are examples of formative methods that address this need, focusing on how products are used in context, and how an understanding of situated use can inform the design of future products. Neither is a silver bullet, but each contributes to the agenda Cockton outlines.

### 15.12.1 References


- Adams, A., Blandford, A. & Lunt, P. (2005) Social empowerment and exclusion: A case study on digital libraries. In *ACM Transactions on CHI*. 12.2. 174-200. DOI http://doi.acm.org/10.1145/1067860.1067863
- Blandford, A., Adams, A., Attfield, S., Buchanan, G., Gow, J., Makri, S., Rimmer, J. & Warwick, C. (2008a) The PRET A Rapporter framework: Evaluating digital libraries from the perspective of information work. *Information Processing & Management*. 44. 4-21. DOI http://dx.doi.org/10.1016/j.ipm.2007.01.021
- Blandford, A. & Furniss, D. (2006) DiCoT: a methodology for applying Distributed Cognition to the design of team working systems. In S. Gilroy & M. Harrison (Eds.) *Proc. DSVIS 2005*. Springer: LNCS 3941. 26-38. DOI http://dx.doi.org/10.1007/11752707_3
- Blandford, A., Green, T. R. G., Furniss, D. & Makri, S. (2008b) Evaluating system utility and conceptual fit using CASSM. *International Journal of Human—Computer Studies*. 66. 393-409. DOI http://dx.doi.org/10.1016/j.ijhcs.2007.11.005
- Blandford, A., Hyde, J. K., Green, T. R. G. & Connell, I. (2008c) Scoping Analytical Usability Evaluation Methods: A Case Study.

*Human Computer Interaction Journal*. 23.3. 278 — 327. DOI 10.1080/07370020802278254

- Blandford, A. & Wong, W. (2004) Situation Awareness in Emergency Medical Dispatch. *International Journal of Human—Computer Studies*. 61(4). 421-452. DOI http://dx.doi.org/10.1016/j.ijhcs.2003.12.012
- Blandford, A. E., Wong, B. L. W., Connell, I. & Green, T. R. G. (2002) Multiple Viewpoints On Computer Supported Team Work: A Case Study On Ambulance Dispatch. In X. Faulkner, J. Finlay & F. Détienne (Eds.) *People and Computers XVI: Proc. HCI'02*. 139-156. Springer./li>
- Carroll, J. M. & Rosson, M.B. (1992) Getting around the task-artifact cycle: how to make claims and design by scenario. *ACM Transactions on Information Systems*, 10(2), 181-212.
- Furniss, D. & Blandford, A. (2006), Understanding Emergency Medical Dispatch in terms of Distributed Cognition: a case study. *Ergonomics Journal*. 49. 12/13. 1174-1203. DOI http://dx.doi.org/10.1080/00140130600612663
- Green, T.R.G. (1989) Cognitive dimensions of notations. In R. Winder and A. Sutcliffe (Eds), *People and Computers V*. Cambridge University Press< 443-460.
- Green, T.R.G., Blandford, A.E., CHurch, L., Roast, C.R. & Clarke, S. (2006) Cognitive dimensions: Achievements, new directions, and open questions, *Journal of Visual Languages & Computing*, Volume 17, Issue 4, Ten Years of Cognitive Dimensions, August 2006, Pages 328-365.
- Hertzum, M., and Jacobsen, N.E. (2001) The Evaluator Effect: A Chilling Fact about Usability Evaluation Methods. *International Journal of Human-Computer Interaction*, 13(4), 421-443.

- John, B. & Kieras, D.E. (1996) Using GOMS for user interface design and evaluation: which technique? *ACM ToCHI* 3.4. 287-319.
- McKnight, J. and Doherty, G. (2008) Distributed cognition and mobile healthcare work. In *Proceedings of the 22nd British HCI Group Annual Conference on People and Computers: Culture, Creativity, Interaction - Volume 2* (BCS-HCI '08), Vol. 2. British Computer Society, Swindon, UK, 35-38.
- Minocha, S., Dawson, L., Blandford, A. & Millard, N. (2005) Providing value to customers in e-commerce environments: the customer's perspective. In *Contemporary Research in E-Marketing Vol 2*. Ideas Group publishing. 119-146.
- Rajkomar, A. & Blandford, A. (in press) Understanding Infusion Administration in the ICU through Distributed Cognition. *Journal of Biomedical Informatics*. To appear.DOI http://dx.doi.org/10.1016/j.jbi.2012.02.003
- Wixon, D. (2003) Evaluating Usability Methods; Why the Current Literature Fails the Practitioner. *Interactions* July and August 2003. 28-34.

## 15.13 COMMENTARY BY THOMAS VISBY SNITKER

How to cite this commentary in your report

### Thomas Visby Snitker

© *Thomas Visby Snitker*

Thomas is one of Denmark's leading usability specialists. He has extensive experience in user research and usability studies for Danish as well as for international companies. Thomas is a frequent speaker at Danish and international conferences, such as The UX Masterclass, and blogs for the Danish edition of Computerworld. In addition, he serves as external reviewer at the IT Universi...

Thomas Visby Snitker
Thomas Visby Snitker is a member of The Interaction Design Foundation

### 15.13.1 Making usability simpler - the way forward?

I work with usability on a daily basis and my clients - most annoyingly - do not really take much interest in what I do for them. Unless of course I break something in the process. Usually they just what to know how they can improve their user interfaces (UI). Well, that's acceptable for me - as a usability specialist that is what I am concerned with.

My customers may ask when to do what and why, but they only listen for as long as it takes to make up their minds — they look for the immediate UI tweaks and solutions, not for insight into the complex intricacies and interactions between users, contexts, media and services. They request my complex research but would rather get a quick fix.

As a consequence, my company has launched a new service, UsabilityForce, to take the complexity out of usability research from the perspective of the customer. UsabilityForce allows producers, designers, developers and others to simply order videos of users thinking aloud while using the clients' service or product at their own leisure, following a test script with test tasks provided by the client. The client can watch the videos and sum up the findings himself or we can provide that through one of our consultants.

The testers install a bit of code on their computer that allows them to hit Record, Pause and Stop. They also use a microphone to capture their audio and an internet connection to upload their video. In the standard test setup with five users, it usually takes only 3-4 hours to collect the five videos.

The contrast between simple and complex research is strong. Quicker versus slower, cheaper versus more expensive, simple versus complex. The ramifications of simpler research are far-reaching and include:

- Less data takes less time to analyse, to report and react to.
- Less time spent in research means quicker reaction time.
- You go into detail with a smaller scope or a specific part of a larger scope.
- Those who find it difficult to see the gains in complex usability because of the size of the investment will have less reason to hesitate. Maybe no reason at all.

- Those who find that complex research takes too much time will find that the simple research is fast.
- Simple testing may make new research topics relevant and feasible; things that where previously too small to test (in terms of costs versus benefits), like a Facebook page, a newsletter, a search engine result page, etc.
- Simple testing can make continuous evaluations feasible - by making simple tests at a regular interval. This could feed into a structured measurement process using Key Performance Indicators (KPIs) all through a development process and ongoing after launch.

I imagine that the simple usability testing will provide a useful supplement to complex testing. As long as complex products and services are conceived and developed they of course need complex research. Furthermore, I speculate that these simpler research technologies will not only have an impact on how usability specialists conduct research, but also on how my 'annoyingly usability ignorant clients' will change. I imagine that some of them will understand better how to benefit from a user research project. They will do the math, they will build the business case, they will include their stake holders and they will persuade reluctant gate keepers.

I also imagine, and hope, that simpler research will allow our community to grow. Those clients who consume our services will grow more committed to the usability of their products and services; they will be more demanding and assertive in the field of usability and perhaps user experience (UX) as well. They will start to question our expertise and they will link their success to the usability of their user interfaces. They will set up a strategy for the user experience (UX targets) and measure the performance of their UX. Simpler research is a strong force and can change how we work and how well usability is adopted by those who need it.

## 15.14 COMMENTARY BY TILDE BEKKER

How to cite this commentary in your report

### Tilde Bekker

Tilde Bekker is an associate professor in the Industrial Design department at the Eindhoven University of Technology. Her research interests include designing for playful interaction, and designing products for children and older adults. She leads and participates in research projects on playful interactions that examine how to persuade people to a healthier lifestyle. She has over 75 pu...

Tilde Bekker
Tilde Bekker is a member of The Interaction Design Foundation

The area of usability evaluation is on the move, as Gilbert Cockton describes. The chapter provides a thorough description of the historical development of usability evaluation methods and provides a good starting point for considering what needs to be done next.

In my commentary I expand on one aspect of evaluation methods: eliciting information from users. I describe how, in the area of *Interaction Design and Children,* evaluation methods have been adapted to increase the output of children participants in evaluation sessions. Two approaches have been applied: by providing different strategies for supporting verbalizations and by providing non-verbal ways to children for expressing their opinion.

For more than 10 years I have been teaching HCI and Industrial Design students how to apply a wide variety of evaluation approaches to various kinds of products and interfaces. Applying evaluation methods to the design of technologies for children can provide a new perspective because it forces us to re-examine some of the assumptions we make about usability evaluation methods.

## 15.14.1 Adapting evaluation methods to participants' skills

What is an interesting challenge in designing and evaluating interactive products for children is to find a good match between the skills and qualities of the participants and the properties of the design and evaluation activity. This approach, which has been widespread in the research area of Interaction Design and Children, has led to some interesting adaptations to existing usability evaluation methods and also to the development of new usability evaluation methods.

In the past 10 to 15 years various studies have examined whether children have the skills and qualities required for a variety of evaluation methods. We can of course argue that when a participant has trouble participating in an evaluation session, we have to train the participant. Another or complementary option is to adjust or redesign the evaluation method to make it easier and possibly more fun to participate in an evaluation session.

## 15.14.2 Verbalization techniques

An important skill required for many evaluation methods is the ability to verbalize one's thoughts. Such verbalizations can be used as a basis for interpreting what usability problems are embedded in the user interface. Different techniques are applied for eliciting verbal output.

One common approach for eliciting verbal output is the *think-aloud method*. Participants are asked to verbalize their thoughts while they are interacting with the product. The evaluation facilitator may prompt the participant to keep talking during the session. However, can children think aloud during usability evaluation sessions? Initially it was suggested that children of 13 years and older can think aloud (Hanna et al., 1997). More recent research showed that children of 7 years and older can think aloud when the protocol for facilitating the verbalizations is adjusted to a more relaxed dialogue (Donker and Markopoulos, 2002).

Evaluation methods may also incorporate other strategies to support participants in verbalizing their thoughts than being prompted by a facilitator. Examples of other strategies are participating in an evaluation session together with peers, tutoring another child, or being prompted by a social robot as a proxy for the facilitator. However, the success of these strategies may depend on children having other skills required for these set-ups, such as the ability to collaborate.

An evaluation method called *co-discovery or constructive interaction* applies a technique where two participants collaborate in performing tasks in an evaluation setting. Supporting verbalizations by talking to a peer may be a more natural setting than holding a monologue or talking to a test facilitator. However, children do need to collaborate for the evaluation sessions to be effective. Some research has shown that younger children of 6 to 7 years old participating may still lack sufficient social skills to be effective participants. They may forget to collaborate and work on a task on their own, thus not providing many verbal utterances. They may sometimes actually compete when doing a task (Markopoulos and Bekker, 2003; Van Kesteren et al., 2003). Older children (between 13 and 14) have been shown to collaborate quite well in co-discovery sessions (Als et al, 2005). Other factors that may influence the quality of the collaboration and the outcome of the session are whether the pairs are friends or not and gender.

Another method, called *peer tutoring*, is based on the idea that one child explains to another child how a product works (Höysniemi et al, 2003). At the beginning one child will try out using a product. Then the first child will become the tutor of a second child. The tutor will help the second child to interact with the product. From the dialogue between the two children their understanding about the product and usability problems can be distilled. The success of this approach depends on whether the tutor is able to fulfill his tutor role effectively, and whether the tutee is open to being taught by another child. Evidence from peer tutoring indicates that when the tutor forgets to play the tutor role the pairs of children take on roles more similar to those in co-discovery sessions. Furthermore, tutors may have trouble only explaining the interaction to the other child without taking over doing a task (Van Kesteren et al, 2003).

A more recently developed method, in which a child is being prompted by a facilitator through a robot interface, is called the *robotic intervention method* (Fransen and Markopoulos, 2010). Providing a context in which children can talk to a playful and toy-like robot is expected to be less inhibiting than talking to an adult. So far, no increase in problems uncovered using this method compared to an active intervention method was found. Children did seem more at ease when participating in the sessions. A slight drawback of the methods was that children perceived the questions asked by the robot to be more difficult than those asked by a human facilitator.

### 15.14.3 Complementing verbal with non-verbal approaches

A different strategy than facilitating verbal output is to provide alternative non-verbal ways to indicate positive and negative aspects in an interface. This was applied in the PhD work by Wolmet Barendregt who developed the *picture cards method* (Barendregt et al., 2008). The method was developed to find problems with children's computer games. It includes cards that children can pick up to indicate both positive and negative aspects in an interface. They can place a card in a box every time they experience a particular emotion shown on one of the cards. They pick up a picture card to express their feelings when interacting with a game or product.

The categories of the cards correspond to various types of problems and fun issues. In a study with children of 5 and 6 years old children expressed more problems explicitly with the picture cards method that in a think-aloud session.

### 15.14.4 A rich usability evaluation context

I agree with Cockton that there are no generalizable evaluation methods. Learning how to conduct usability evaluations requires developing an understanding of the complete evaluation context. This context includes many factors, such as who applies the method in what type of development process. And it also includes, as I illustrated earlier, specific requirements of the user group.

Evaluation methods can be further improved by adapting them to the skills and qualities of all the stakeholders involved, by providing diverse ways to provide input, addressing positive and negative experiences, and possibly even making the activity more fun and enjoyable.

Developing evaluation approaches is like developing products and systems: for every improvement we try to incorporate in an evaluation method, we run the risk of adding new challenges for the participants.


Figure 15.1: Usability Evaluation with children.



## 15.14.5 References


- Als, B.S., Jensen, J.J. & Skov, M.B. (2005) Exploring verbalization and collaboration of constructive interaction with children. In *Proceedings of the 2005 IFIP TC13 international conference on Human-Computer Interaction* (INTERACT'05), Springer-Verlag, Berlin, Heidelberg, 443-456.
- Barendregt , W., Bekker , M.M., & Baauw, E. (2008) Development and evaluation of the problem identification picture cards method, Cognition, Technology and Work, 10 (2), 95-105.

Donker, A., & Markopoulos, P. (2002) A comparison of think-aloud, questionnaires and interviews for testing usability with children. In *People and Computers XVI - Memorable Yet Invisible, Proceedings of HCI 2002*, London: Springer, 305-316.

- Fransen, S. & Markopoulos, P. (2010). Let robots do the talking. In *Proceedings of the 9th International Conference on Interaction Design and Children* (IDC '10). ACM, New York, NY, USA, 59-68. Hanna ,J., Risden , K. & Alexander, K. Guidelines for usability testing with children, (1997) interactions, 4(5), 9-14.
- Höysniemi,J., Hämäläinen, P. &Turkki, L. (2003) Using peer tutoring in evaluating the usability of a physically interactive computer game with children. Interacting with Computers 15(2): 203-225
- Markopoulos, P. & Bekker, M. (2003) On the assessment of usability testing methods for children. *Interacting with Computers*, 15(3), 227-243.
- Van Kesteren, I., Bekker, M.M., Vermeeren, A.P.O.S. & Lloyd, P. (2003) Assessing usability evaluation methods on their effectiveness to elicit verbal information from children subjects, In *Proceedings of the Conference on Interaction Design and Children*, Preston, 1-3 July 2003, 41-49.

## 15.15 REFERENCES

*Bardzell, Jeffrey (2011): Interaction criticism: An introduction to the practice. In Interacting with Computers, 23 (6) pp. 604-621*

*Bardzell, Jeffrey (2009): Interaction criticism and aesthetics. In: Proceedings of ACM CHI 2009 Conference on Human Factors in Computing Systems 2009. pp. 2357-2366*

*Bardzell, Shaowen and Bardzell, Jeffrey (2011): Towards a feminist HCI methodology: social science, feminism, and HCI. In: Proceedings of ACM CHI 2011 Conference on Human Factors in Computing Systems 2011. pp. 675-684*

*Bellamy, Rachel, John, Bonnie E. and Kogan, Sandra (2011): Deploying CogTool: integrating quantitative usability assessment into real-world software development. In: Proceeding of the 33rd international conference on Software engineering 2011. pp. 691-700*

*Blythe, Mark, Petrie, Helen and Clark, John A. (2011): F for fake: four studies on how we fall for phish. In: Proceedings of ACM CHI 2011 Conference on Human Factors in Computing Systems 2011. pp. 3469-3478*

*Brown, Tim (2009): Change by Design: How Design Thinking Transforms Organizations and Inspires Innovation. HarperBusiness*

*Brown, Barry A. T., Taylor, Alex S., Izadi, Shahram, Sellen, Abigail, Kaye, Joseph Jofish and Eardley, Rachel (2007): Locating Family Values: A Field Trial of the Whereabouts Clock. In: Krumm, John, Abowd, Gregory D., Seneviratne, Aruna and Strang, Thomas (eds.) UbiComp 2007 Ubiquitous Computing - 9th International Conference September 16-19, 2007, Innsbruck, Austria. pp. 354-371*

*Card, Stuart K., Moran, Thomas P. and Newell, Allen (1980): The keystroke-level model for user performance with interactive systems. In Communications of the ACM, 23 pp. 396-410*

*Cockton, Gilbert (2006): Designing worth is worth designing. In: Proceedings of the Fourth Nordic Conference on Human-Computer Interaction 2006. pp. 165-174*

*Cockton, Gilbert (2007): Some Experience! Some Evolution. In: Erickson, Thomas and McDonald, David W. (eds.). "HCI Remixed: Reflections on Works That Have Influenced the HCI Community". The MIT Presspp. 215-219*

*Cockton, Gilbert (2004): Value-centred HCI. In: Proceedings of the Third Nordic Conference on Human-Computer Interaction October 23-27, 2004, Tampere, Finland. pp. 149-160*

*Cockton, Gilbert (2008): Load while Aiming; Hit? (invited keynote). In: Brau, Henning (ed.). "Usability Professionals 2008 Conference". pp. 17-22*

*Cockton, Gilbert and Lavery, Darryn (1999): A Framework for Usability Problem Extraction. In: Angela.Sasse, M. and Johnson, Chris (eds.) Proceedings of Interact 99 1999, Edinburgh.*

*Cockton, Gilbert and Woolrych, Alan (2009): A. Comparing UEMs: Strategies and Implementation, Final Report of COST 294 Working Group 2. In: Law, Effie Lai-Chong, Scapin, Dominique, Cockton, Gilbert, Springett, Mark, Stary, C. and Winckler, Marco (eds.). "Maturation of Usability Evaluation Methods: Retrospect and Prospect: Final Reports of COST294-MAUSE Working Groups".*

*Cockton, Gilbert, Woolrych, Alan, Hornbæk, Kasper and Frøkjær, Erik (2012): Inspection-based methods. In: Sears, Andrew and Jacko, Julie A. (eds.). "The Human-Computer Inter-*

*action Handbook: Fundamentals, Evolving Technologies and Emerging Applications, Third Edition". pp. 1275-1293*

*Cockton, Gilbert, Woolrych, Alan and Hindmarch, Mark (2004): Reconditioned merchandise: extended structured report formats in usability inspection. In: Dykstra-Erickson, Elizabeth and Tscheligi, Manfred (eds.) Extended abstracts of the 2004 Conference on Human Factors in Computing Systems, CHI 2004, Vienna, Austria, April 24 - 29, 2004 2004. pp. 1433-1436*

*Cockton, Gilbert, Kujala, Sari, Nurkka, Piia and Hölttä, Taneli (2009): Supporting Worth Mapping with Sentence Completion. In: Gross, Tom, Gulliksen, Jan, Kotze, Paula, Oestreicher, Lars, Palanque, Philippe A., Prates, Raquel Oliveira and Winckler, Marco (eds.) Human-Computer Interaction - INTERACT 2009, 12th IFIP TC 13 International Conference, Uppsala, Sweden, August 24-28, 2009, Proceedings, Part II 2009. pp. 566-581*

*Cockton, Gilbert, Kirk, Dave, Sellen, Abigail and Banks, Richard (2009): Evolving and augmenting worth mapping for family archives. In: Proceedings of the HCI09 Conference on People and Computers XXIII 2009. pp. 329-338*

*Dumas, Joseph S. and Fox, Jean E. (2007): Usability Testing: Current Practice and Future Directions. In: Sears, Andrew and Jacko, Julie A. (eds.). "The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications, Second Edition (Human Factors and Ergonomics)". CRC Presspp. 1129-1150*

*Dumas, Joseph S. and Redish, Janice C. (1993): A Practical Guide to Usability Testing. Norwood, NJ, Intellect*

*Erickson, Thomas and McDonald, David W. (2008): HCI Remixed: Reflections on Works that Have Influenced the HCI Community. Cambridge, Massachusetts, MIT Press*

*Forlizzi, Jodi (2008): The product ecology: Understanding social product use and supporting design culture. In International Journal of Design, 2 (1) pp. 11-20*

*Gaver, William W., Boucher, Andy, Law, Andy, Pennington, Sarah, Bowers, John, Beaver, Jacob, Humble, Jan, Kerridge, Tobie, Villar, Nicholas and Wilkie, Alex (2008): Threshold devices: looking out from the home. In: Proceedings of ACM CHI 2008 Conference on Human Factors in Computing Systems April 5-10, 2008. pp. 1429-1438*

*Gray, Wayne D. and Salzman, Marilyn C. (1998): Damaged Merchandise? A Review of Experiments That Compare Usability Evaluation Methods. In Human-Computer Interaction, 13 (3) pp. 203-261*

*Harper, Richard, Rodden, Tom, Rogers, Yvonne and Sellen, Abigail (2008): Being Human: Human Computer Interaction in 2020. Microsoft Research Ltd*

*Hertzum, Morten and Jacobsen, Niels Ebbe (2001): The Evaluator Effect: A Chilling Fact About Usability Evaluation Methods. In International Journal of Human-Computer Interaction, 13 (4) pp. 421-443*

*Herzberg, Frederick (1966): Work and the Nature of Man.*

*Herzberg, Frederick (1973): Work and the Nature of Man. Signet*

*Hornbæk, Kasper (2010): Dogmas in the assessment of usability evaluation methods. In Behaviour and Information Technology, 29 (1) pp. 97-111*

*Iivari, N. (2005): Usability Specialists - 'Mommy Mob', 'Realistic Humanists' or 'Staid Researchers'? An Analysis of Usability Work in the Software Product Development. In: Proceedings of IFIP INTERACT05: Human-Computer Interaction 2005. pp. 418-430*

*International Standards Association (1991). ISO/IEC 9126: Software engineering - Product quality. Retrieved 1 December 2011 from International Standards Association: http://en.wikipedia.org/wiki/ISO/IEC_9126*

*International Standards Association (1998). ISO 9241-11: Ergonomic requirements for office work with visual display terminals (VDTs) -- Part 11: Guidance on usability. Retrieved 1 December 2011 from International Standards Association: http://en.wikipedia.org/wiki/ISO/IEC_9126*

*International Standards Association (2001). ISO/IEC 9126-1:2001 Software engineering - Product quality - Part 1: Quality model,. Retrieved 1 December 2011 from International Standards Association: http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=22749*

*International Standards Association (2004). ISO/IEC TR 9126-4:2004 Software engineering -- Product quality -- Part 4: Quality in use metrics. Retrieved 1 December 2011 from International Standards Association: http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=39752*

*International Standards Association (2011). ISO/IEC 25010 Systems and software engineering -- Systems and software Quality Requirements and Evaluation (SQuaRE) -- System and software quality models. Retrieved 1 December 2011 from International Standards Association: http://www.iso.org/iso/catalogue_detail.htm?csnumber=35733*

*John, Bonnie E. and Kieras, David E. (1996): Using GOMS for User Interface Design and Evaluation: Which Technique?. In ACM Transactions on Computer-Human Interaction, 3 (4) pp. 287-319*

Landauer, Thomas K. (1996): The Trouble with Computers: Usefulness, Usability, and Productivity. The MIT Press

Lavery, Darryn, Cockton, Gilbert and Atklnson, Malcolm P. (1997): Comparison of Evaluation Methods Using Structured Usability Problem Reports. In Behaviour and Information Technology, 16 (4) pp. 246-266

Law, Effie Lai-Chong, Hvanberg, Ebba, Cockton, Gilbert, Palanque, Philippe A., Scapin, Dominique L., Springett, Mark, Stary, Christian and Vanderdonckt, Jean M. (2005): Towards the maturation of IT usability evaluation (MAUSE). In: Costabile, Maria Francesca and Paterno, Fabio (eds.). "Human-computer interaction: INTERACT 2005: IFIP TC13 international conference: proceedings. Lecture Notes in Computer Science. (3585)". pp. 1134-1137

Nielsen, Jakob (1994): Enhancing the explanatory power of usability heuristics. In: Plaisant, Catherine (ed.) Conference on Human Factors in Computing Systems, CHI 1994, Boston, Massachusetts, USA, April 24-28, 1994, Conference Companion 1994. p. 210

Otero, Nuno and José, Rui (2009): Worth and Human Values at the Centre of Designing Situated Digital Public Displays. In IJAPUC, 1 (4) pp. 1-13

Pew, Richard W. (2002): Evolution of human-computer interaction: from Memex to Bluetooth and beyond. In: Sears, Andrew and Jacko, Julie A. (eds.). "The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications (Human Factors and Ergonomics)". Lawrence Erlbaumpp. 1-17

Rodden, Kerry, Hutchinson, Hilary and Fu, Xin (2010): Measuring the user experience on a large scale: user-centered metrics for web applications. In: Proceedings of ACM CHI 2010 Conference on Human Factors in Computing Systems 2010. pp. 2395-2398

Rogers, Yvonne, Bannon, Liam and Button, Graham (1994): Rethinking theoretical frameworks for HCI: A Review. In ACM SIGCHI Bulletin, 26 (1) pp. 28-30

Rosenbaum, Stephanie (2008): The Future of Usability Evaluation : Increasing Impact on Value. In: Law, Effie Lai-Chong, Hvannberg, Ebba Thora and Cockton, Gilbert (eds.). "Maturing Usability". Springerpp. 344-378

Rosenbaum, Stephanie, Rohn, Janice Anne and Humburg, Judee (2000): A Toolkit for Strategic Usability: Results from Workshops, Panels, and Surveys. In: Turner, Thea, Szwillus, Gerd, Czerwinski, Mary, Peterno, Fabio and Pemberton, Steven (eds.) Proceedings of the ACM CHI 2000 Human Factors in Computing Systems Conference April 1-6, 2000, The Hague, The Netherlands. pp. 337-344

*Salvucci, Dario D. (2009): Rapid prototyping and evaluation of in-vehicle interfaces. In ACM Transactions on Computer-Human Interaction, 16 (2) p. 9*

*Sears, Andrew and Jacko, Julie A. (2007): The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications (2nd Edition). Lawrence Erlbaum Associates*

*Sengers, Phoebe and Gaver, William (2006): Staying open to interpretation: engaging multiple meanings in design and evaluation. In: Proceedings of DIS06: Designing Interactive Systems: Processes, Practices, Methods, & Techniques 2006. pp. 99-108*

*Shneiderman, Ben (1986): No members, no officers, no dues: A ten year history of the software psychology society. In ACM SIGCHI Bulletin, 18 (2) pp. 14-16*

*Smith, Sidney L. and Mosier, Jane N. (1986). GUIDELINES FOR DESIGNING USER INTERFACE SOFTWARE, The MITRE Corporation, Bedford, Massachusetts, USA, Prepared for Deputy Commander for Development Plans, and Support Systems, Electronic Systems Division, AFSC, United States Air Force, Hanscom Air Force B.*

*Venturi, Giorgio, Troost, Jimmy and Jokela, Timo (2006): People, Organizations, and Processes: An Inquiry into the Adoption of User-Centered Design in Industry. In International Journal of Human-Computer Interaction, 21 (2) pp. 219-238*

*Wasserman, Anthony I. (1973): The design of 'idiot-proof' interactive programs. In: Proceedings of the June 4-8, 1973, national computer conference and exposition 1973. pp. m34-m38*

*Whiteside, John, Bennett, John and Holtzblatt, Karen (1988): Usability Engineering: Our experience and Evolution. In: Helander, Martin and Prabhu, Prasad V. (eds.). "Handbook of human-computer interactio". pp. 791-817*

*Wilson, Chauncey (1999). Severity Scale for Classifying Usability Problems. Retrieved 1 December 2011 from The Usability SIG Newsletter: http://www.stcsig.org/Usability/newsletter/9904-severity-scale.html*

*Woolrych, Alan, Hornbæk, Kasper, Frøkjær, Erik and Cockton, Gilbert (2011): Ingredients and Meals Rather Than Recipes: a Proposal for Research That Does Not Treat Usability Evaluation Methods As Indivisible Wholes. In International Journal of Human-Computer Interaction, 99999 (1) pp. 1-1*

## About the author

### Gilbert Cockton

Has also published under the name of:
"G. Cockton"

Personal Homepage:
designsituations.blogspot.com/

Current place of employment:
Northumbria University

I'm Professor of Design Theory in the School of Design at Northumbria University, which has roots back to 1844 as one of the original British Government Schools of Design. I'm an Interaction Designer who occasionally dabbles in product and service design. I have a multi-disciplinary background in humanities (History), applied human sciences (Education), and engineering design (Computer Science, old school HCI). In 2005 I was awarded a UK NESTA fellowship to work on value-centred approaches to design. As a first result, I moved beyond value (to worth) and beyond centredness (to multiple design foci). As a second result, my research has developed a more general design focus, and I moved from Computing (Sunderland) to Design (Northumbria) in September 2009. I now work alongside very talented design educators and researchers with a broad range of craft skills and design philosophies. I increasingly find the user-centred positions of 1980s HCI naive and uninformed. When I'm not supporting an amazing group of colleagues in my role as Associate Dean for Research and Innovation, my research focuses on balanced, integrated and generous (BIG!) fusions of the main design paradigms (Applied Arts, Engineering, User-Centred). Examples of such fusions include leveraging crafted forms within user research, focusing evaluation practices on achieved worth, combining engineering specification with humane design purposes, and blending tacit creative and explicit systematic design work. I have developed the Working to Choose (W2C) framework to provide an overarching structure for co-ordinating research on re-usable resources in design practice.

## Your notes and thoughts on chapter 15

Record your notes and thoughts on this chapter. If you want to share these thoughts with others online, go to the bottom of the page at:
http://www.interaction-design.org/encyclopedia/usability_evaluation.html

**NOTES:**

# Chapter 16

# Activity Theory

*by Victor Kaptelinin.*

Foreword: Why activity theory?

This chapter is about a theory that was developed decades ago. Some of the basic ideas of the theory were formulated before the word "computer" was ever invented. Then why does the Encyclopaedia of Human-Computer Interaction feature a chapter on the theory? In other words, *Why activity theory?*

The question can be answered in two steps.

(a) Why *activity*?

Activity is currently one of the most fundamental concepts in HCI research (Moran, 2006). Early HCI was predominantly concerned with understanding and supporting *tasks*, which people do to achieve clear predetermined goals (such as making certain changes in a document). The issues of *why* a person carries out a task and what the task *means* to the person were typically outside the scope of analysis, evaluation, and design. However, with interactive technology becoming a part of our everyday environments the focus on tasks proved to be insufficient. Understanding and designing technology in the context of purposeful, meaningful *activities* is now a central concern of HCI research and practice. Virtually all significant recent developments in interactive technologies — think about, for instance, social media, smartphones, and bookreaders — owe their success to helping us live fuller lives rather than merely supporting new types of tasks.

(b) Why activity *theory*?

Most people have an intuitive understanding of what activities are. Is there any need for a theory here?

A problem with intuitive, commonsense notions of activity is that they can be different for different people. In addition, they may be not specific enough. How to distinguish activities from non-activities? Can activities be broken down into smaller units? What role does technology play in human activity? To answer these and other similar questions HCI needs a more elaborated concept of activity. Such concept is offered by activity theory, discussed in this chapter.

## 16.1 INTRODUCTION

Activity theory is a conceptual framework originating from the socio-cultural tradition in Russian psychology. The foundational concept of the framework is "activity", which is understood as purposeful, transformative, and developing interaction between actors ("subjects") and the world ("objects"). The framework was originally developed by the Russian psychologist Aleksei Leontiev[1] (Leontiev 1978; Leontiev 1981). A version of activity theory, based on Leontiev's framework,

1. The Russian last name "Леонтьев" is variously spelled in Latin alphabet as "Leontiev", "Leontev", "Leont'ev", "Leontyev", etc. To avoid confusion, the present chapter only uses one spelling, "Leontiev".

was proposed in the 1980s by the Finnish educational researcher Yrjö Engeström (1987). Currently, both Leontiev's and Engeström's variants of activity theory, as well as their combinations, are being widely used interdisciplinarily, not only in psychology, but also in a range of other fields, including education, organizational learning, and cultural studies.

Since the early 1990s, activity theory has been a visible landmark of the theoretical landscape of Human-Computer Interaction (HCI). In the last two decades, activity theory, along with some other frameworks, such as distributed cognition and phenomenology, has established itself as a leading post-cognitivist approach in HCI and interaction design (e.g., Bødker, 1991; Nardi, 1996a; Bertelsen and Bødker, 2003; Kaptelinin et al., 2003; Kaptelinin and Nardi, 2006). Carroll 2011 observes that: "Information processing psychology and laboratory user studies, once the kernel of HCI research, became important, but niche areas. The most canonical theory-base in HCI now is socio-cultural, Activity Theory."

This chapter discusses the past, present, and future of activity theory as a theoretical approach in HCI. It starts with a brief introduction to the basic concepts and principles of activity theory, continues to describe its key contributions to research in HCI and interaction design, and concludes with reflections on challenges and prospects for further development of the approach.

The chapter is not intended to be a comprehensive exposition of the framework and its uses in HCI. More detailed discussions of activity theory concepts and applications in the context of HCI research can be found for instance, in Bødker (1991), Nardi (1996a), Engeström et al. (1999), Kaptelinin and Nardi (2006), and Kaptelinin and Nardi (2012))

## 16.2 BRIEF OVERVIEW OF ACTIVITY THEORY

### 16.2.1 Historical roots and underlying assumptions

The immediate conceptual roots of activity theory can be traced to Russian/Soviet psychology of the 1920s and 1930s.[2] During that time theoretical explorations in Russian psychology were heavily influenced by Marxist philosophy. A collective effort of a number of prominent psychologists, most notably Lev Vygotsky and Sergey Rubinshtein—which effort also involved much disagreement and even open conflicts—gave rise to a socio-cultural perspective (understood in a broad sense) in Russian psychology (e.g., Vygotsky, 1978; Rubinshtein, 1946; Rubinshtein 1986).

The main conceptual thrust of the socio-cultural perspective was to overcome the divide between, on the one hand, human mind, and on the other hand, culture and society. As opposed to most psychological frameworks of that time, the perspective considered culture and society generative forces, "responsible" for the very production of human mind, rather than external factors, however important, that merely constitute conditions for the functioning of the mind without changing its basic nature.

The work based on the socio-cultural perspective produced a number of fundamental insights. Some of the most important contributions were as follows:

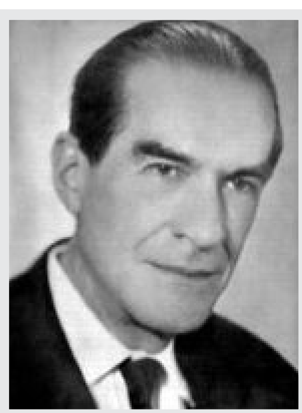
- Vygotsky's universal *law of development*, according to which human mental functions first emerge as distributed between the person and other people (i.e., "inter-psychological ") and only then as individually mastered by the person himself or herself (i.e., "intra-psychological"), and
- Rubinshtein's principle of "*unity and inseparability of consciousness and activity*", according to which human conscious experi-

2. More broadly, activity theory represents an intellectual tradition that has been manifested throughout ages in a variety of seemingly diverse schools of thought which one way or another emphasize the generative and transformative nature of purposeful human action. The tradition can be traced, for instance, to Hegel, Goethe's philosophical poetry, and even Buddhism

> ence and human acting in the world, the internal and the external, are closely interconnected and mutually determine one another.

Aleksei Leontiev's activity theory[3] emerged as an outgrowth of the socio-cultural perspective. The theory employs a number of ideas developed by Vygotsky, Leontiev's mentor and friend. It is also strongly influenced by the work of Rubinshtein, a major figure in Russian psychology and a long-time colleague of Leontiev's (Brushlinsky and Aboulhanova-Slavskaya, 2000). Arguably, activity theory also features some other important influences which are more difficult to discern, such as the framework developed by Mikhail Basov (Basov, 1991). The basic assumptions of activity theory are the same as those underlying the socio-cultural perspective in general: namely, the assumptions of *the social nature of human mind* and *inseparability of human mind and activity*.

At the same time, Leontiev's activity theory is not a simple imprint of all these influences. As discussed below, while the theory incorporates a variety of ideas developed by Vygotsky, Rubinshstein, and others, these ideas have been revised and elaborated upon by Leontiev to form his own distinct and consistent conceptual framework.





3. Leontiev himself usually referred to his framework as "the activity approach ("деятельностный подход") in psychology", rather than "activity theory" (cf. Mescherjakov and Zinchenko, 2004)





FIGURE 16.1 A-B-C: From left to right: Aleksei Nikolaevich Leontiev, Lev Semenovich Vygotsky, and Sergey Leonidovich Rubinshtein.

## 16.3 BASIC CONCEPTS AND PRINCIPLES OF LEONTIEV'S FRAMEWORK

### 16.3.1 The concept of 'activity'

Activity, in a broad sense, is an interaction of the actor (e.g., a human being) with the world. The interaction, according to activity theory terminology, is described as a process relating the *subject* (S) and the *object* (O). A common way to represent activity is "S ⇔ O". There are two key aspects differentiating activity from other types of interaction: (a) subjects of activities *have needs*, which should be met through an interaction with the world, and (b) activities and their subjects mutually determine one another; or, more generally, activities are generative forces that *transform both subjects and objects*.

*Subjects have needs*. Activity is understood as a "unit of life" of a material subject existing in the objective world. Subjects have their own needs and, in order to survive, have to carry out activities, that is, interact with objects of the world to meet their needs. Leontiev's analysis was mostly concerned with activities of individual human beings, but the notion of "subject" is not limited to individual humans. Other types of entities, such as animals, teams, and organizations can also have need-based agency and, therefore, be subjects of activities (Kaptelinin and Nardi, 2006).

*Activities and their subjects mutually determine one another*. It is immediately obvious that activities are influenced by the attributes of subjects and objects. Consider a simple example. Undoubtedly, whether or not a person can solve a math problem depends on the nature of the problem (e.g., how difficult it is) and the person's abilities and skills (i.e., how good the person is at math). In the long run, however, the opposite is also true: both the object *and the subject* are over time *transformed by* the activity. It is apparent, for instance, that a person's math skills are a result of previous experience: they have developed through solving math problems in the past. In other words, while it is true that a person's math abilities determine how the person solves math problems, it is also true that solving math problems determine the person's math abilities. Therefore, subjects do not only express themselves in their activities; in a very real sense they are *produced* by the activities (cf. Rubinshtein, 1986).

**Mind and activity: Leontiev vs. Rubinshtein**

Leontiev extends and develops Rubinshtein's principle of unity and inseparability of consciousness and activity in three respects. First, Leontiev states that psychological studies should not be focusing only on the "psychological aspect or facet of activity" (as suggested by Rubinshtein), such as the relationship between activity and subjective experiences. Instead, he maintained that the relevance of activity to psychology is of a more general nature: activity is of fundamental importance to psychology because of its special *function*, the function of placing the subject in the objective reality and transforming this reality into a form of subjectivity (Leontiev, 1978). Second, as discussed below, Leontiev's analysis focuses on both conscious and unconscious mental phenomena. Third and finally, Leontiev offered a number of more concrete insights about the relationship between mind and activity, most notably the idea of structural similarity between internal and external processes (Leontiev, 1978; Leontiev, 1981).

### 16.3.2 Basic principles

The main ideas and assumptions of activity theory, outlined above, have been elaborated by Leontiev into a set of more specific notions, claims, and arguments. A common problem with interpreting Leontiev's texts is that they often reflect the unfolding logic of his conceptual explorations rather than provide a systematic overview of the logical structure of the framework as a whole. There have been several attempts to translate the representation of Leontiev's framework, as it is described in his texts, into a structured set of distinct principles. Kaptelinin and Nardi (2006), building on Wertsch (1981), identify the following principles:

#### 16.3.2.1 Object-orientedness

This principle (which bears some similarity to phenomenology's notion of "intentionality" – see Dourish, 2001) is directly related to the very concept of activity as a "subject-object" relationship. Why is subjects' interaction with the *world* defined in terms of interacting with *objects*? The explanation, offered by the principle of object-orientedness, is as follows. The world is structured; it comprises discrete objectively existing entities, that is, objects. Subjects' interaction with the world is also structured; it is organized around the objects. Objects have their "objective" meanings, determined by their relationship with other entities existing in the world (including the subject). In order to meet their needs, the subject has to reveal the objective meaning of the objects, at least partly, and act accordingly.

Therefore, the object of activity has two facets, it should be understood:

- First, in its independent existence as subordinating to itself and transforming the activity of the subject
- Second, as an image of the object, as a product of its property of psychological reflection that is realized as an activity of the subject and cannot exist otherwise (Leontiev, 1978).

These two facets do not necessarily always coincide. They are dynamically aligned in the unfolding "subject-object" interaction. The alignment involves a double transition: the subject's activity is subordinated to properties of the object which gives rise to new activity structures; in turn, new activity structures bring about new subjective phenomena, such as a more developed image of the object. For instance, a tourist wandering around an area may initially have a vague idea about the area and simply follow the constraints and possibilities provided by the environment. Over time, emerging patterns of walking may result in a development of an elaborated cognitive map of the area.

The principle of object-orientedness applies differently to animals and human beings. Animals live in a structured world of natural objects which are material and mostly have direct positive or negative meanings and values, provide affordances for action, and so forth. Human beings live in a predominantly man-made world, where objects are not necessarily physical things: they can be intangible, but they can still be considered "objects" as long as they objectively exist in the world. For instance, the objects of learning a new language or making a company profitable are impossible to touch, physically weigh, or measure with a ruler. However, the grammatical structure of a language or profit margin of a company does not exist merely in a person's imagination. Rather, they are "facts of life", which need to be faced and dealt with. "Objective" is understood in activity theory in a broad sense as including not only the properties of things that can be directly registered with physical instruments, but also socially and culturally defined properties.

Therefore, the principle of object-orientedness states that all human activities are directed toward their objects and are differentiated from one another by their respective objects. Objects motivate and direct activities, around them activities are coordinated, and in them activities are crystallized when the activities are complete. Analysis of objects is therefore a necessary requirement for understanding human activities, both individual and collective ones.

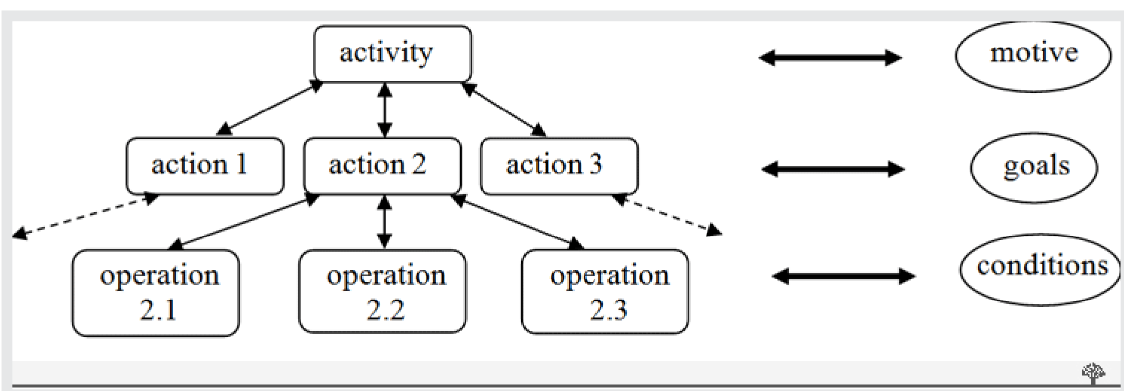



FIGURE 16.2: Hierarchical structure of activity.

## Lost in translation: “Predmet” vs. “objekt”

There is a language problem, which makes an adequate translation of Leontiev’s notion of “object” from Russian to English somewhat complicated. In Russian there are two words which have similar but distinct meanings: “objekt” and “predmet”. Both refer to objectively existing entities, but the notion of “predmet” typically also implies a relevance of the entity in question to certain human purposes or interests.[1] Similar linguistic distinctions can be found in German and some other languages. Leontiev deliberately referred to the object of activity as “predmet” rather than “objekt”. However, this distinction is usually lost in English translation since both words are translated as “object”. This linguistic problem is a likely reason why the emergence of objects of activities—the dynamics of “just objects” becoming involved in activities and acquiring the status of “objects of activities”, and vice versa—have not so far received the attention they deserve in concrete studies informed by activity theory.

1. In this respect the distinction between “objekt” and “predmet” is somewhat similar to the distinction between “space” and “place”, which became popular in the fields of HCI and CSCW (see e.g. Dourish, 2001)

### 16.3.2.2 Hierarchical structure of activity

Human activities, according to Leontiev, are units of life which are organized into three hierarchical layers (see Figure 16.2). The top layer is the *activity* itself, which is oriented toward a *motive*, corresponding to a certain need. The motive is the object that the subject ultimately needs to attain. For instance, in some cultural contexts people reaching a certain age *need* to learn how to drive a car (and get a driver’s license); it is a general prerequisite of being a fully functional member of society. Learning how to drive a car is an activity which is organized as a multi-layer system of sub-units directed at getting a driver's license. *Actions* are conscious processes directed at *goals* which must be undertaken to fulfil the object. Goals can be decomposed into sub-goals, sub-sub-goals, and so forth. For instance, one may decide to enroll in a driving school, purchase instructional materials, make a schedule of theoretical lessons and practice sessions, etc. Actions are implemented through lower-level units of activity, called *operations*. Operations are routine processes providing an adjustment of an action to the ongoing situation. They are oriented toward the *conditions* under which the subject is trying to attain a goal. People are typically not aware of their operations. For instance, a driving school student taking notes during a lecture might be fully concentrated on traffic rules rather than the process of writing. Operations emerge in two ways. First, an operation can be a result of step-by-step automatization of an originally conscious action (e.g., over time, the action of changing lanes may transform into a routine operation, which does not require conscious control). When such operations fail, they are often transformed into conscious actions again. Second, an operation can be a result of "improvisation", a spontaneous adjustment of an action on the fly (e.g., in an emergency situation the driver may act "instinctively", without thinking).

The three-layer model only applies to human activities. Complex relationships between motives (i.e., what motivates the activity) and goals (i.e., what directs the activity) is a characteristic feature of humans. While animals usually act directly toward the objects that motivate them (e.g., food), humans often attain their motives by directing their efforts to other things (e.g., however hungry, people usually grab a menu, rather than the first available food, upon entering a restaurant). This feature, according to Leontiev, is a product of the complex social organization of human life. In particular, the emergence of division of labour entails the need for some people to focus on objects, different from the ones that actually meet a certain need. For instance, the actions of primordial hunters who scare the game away (i.e., "beaters") may look paradoxical if one does not know that the game is directed toward another group of hunters, waiting in the ambush (i.e., “ambushers”). Once the feature of the social organization of life, the dissociation between motivating objects (motives) and directing objects (goals) shapes the structure of individual activities and becomes its characteristic feature, as well (Leontiev, 1981).

Considering human activity as a three-layer system opens up a possibility for a combined analysis of motivational, goal-directed, and operational aspects of human acting in the world, that is, bringing together the issues of Why, What, and How within a consistent conceptual framework (Bødker, 1991). Realizing this possibility in a concrete study may, however, be problematic. Revealing the ultimate motives of a person or the fine-grain structure of automatic operations may prove to be difficult, if not impossible. This limitation of Leontiev’s three-layer model as an analytical tool can be overcome by employing an expansive “actions first” strategy. This strategy involves starting analysis from the actions layer which relatively easily yields itself to qualitative research methods. In particular, people are usually aware of their goals and can report or express them in a certain way. Then the analysis can be expanded both “up”, to progressively higher level goals and, ultimately, motives, and “down”, to sub-goals and operations. The expanding scope of analysis may not cover the entire structure of the activity in question but be sufficient for the purposes of the task at hand (see also Kaptelinin and Nardi, 2006).

#### 16.3.2.3 Mediation

Arguably, mediation is the primary dimension along which human beings differ from other animals. It is mediation which has made *homo sapiens* such a successful species: while we do not have sharp claws and thick fur, we compensate that by employing mediating artefacts, such as hammers, knives, and warm clothes. The effect of complex social organization on the structure of individual human activity, discussed above and illustrated by the case of primordial hunters, is another example of mediation. In fact, the main distinctive features of humans, such as language, society and culture, the production and use of advanced tools, etc., all involve mediation. They represent different aspects of the same phenomenon, that is, the emergence of a complex system of objects and structures, both material and immaterial which serve as mediating means embedded in the interaction between human beings and the world and shaping the interaction.

Activity theory inherits its special interest in mediation from the approach that made the most fundamental impact on Leontiev's framework – that is, Vygotsky's cultural-historical psychology. In cultural-historical psychology mediation is, arguably, *the* most important concept of all; it serves as the cornerstone of the approach as a whole. Vygotsky proposed that the very nature of human mental processes, as opposed to animals' mental processes, is defined by mediation. Vygotsky's ideas concerning mediation were explicitly incorporated into the conceptual framework of activity theory but placed in a somewhat different theoretical context. As opposed to Vygotsky, who was predominantly interested in particular higher mental functions and their ontogenetic development and, therefore, particularly concerned with means that mediate specific mental operations (especially, signs), Leontiev's mainly focussed on means that mediate a purposeful object-oriented activity as a whole.

Tool mediation allows for appropriating socially developed forms of acting in the world. Tools reflect the previous experience of other people, which experience is accumulated in the structural properties of tools, such as their shape or material, as well as in the knowledge of how the tool should be used (see Figure 16.3 to 16.8). Therefore, the use of tools is a form of accumulation and transmission of social, cultural knowledge. Tools not only shape the external behaviour. As discussed below, through internalization they also influence the mental functioning of individuals. For instance, a person's cognitive map of a city may depend on whether or not the person is a car driver. Some folklore sayings also suggest that our perception of the world is affected by the tools we are using, e.g.: "If all you have is a hammer, everything looks like a nail."

### 16.3.2.4 Example of mediation and accumulation of experience over time: Devices for calculation and computation


Figure 16.3: Woodcut of a “Calculating-Table” by Gregor Reisch, 1508. The woodcut shows Arithmetica instructing an algorist and an abacist.

*Courtesy of Gregor Reisch,1508. Copyright: pd (Public Domain (information that is common property and contains no original authorship)).*

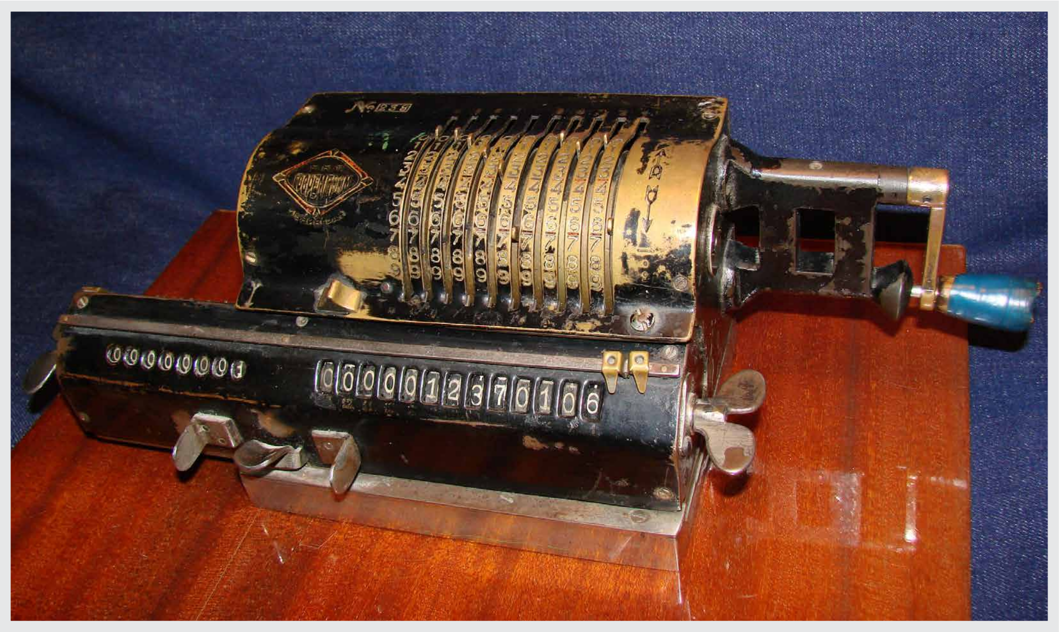


Figure 16.4: Soviet-produced calculator from the Soviet Calculators Collection.

*Courtesy of Sergei Frolov and the Soviet Calculators Collection. Copyright: pd (Public Domain (information that is common property and contains no original authorship)).*

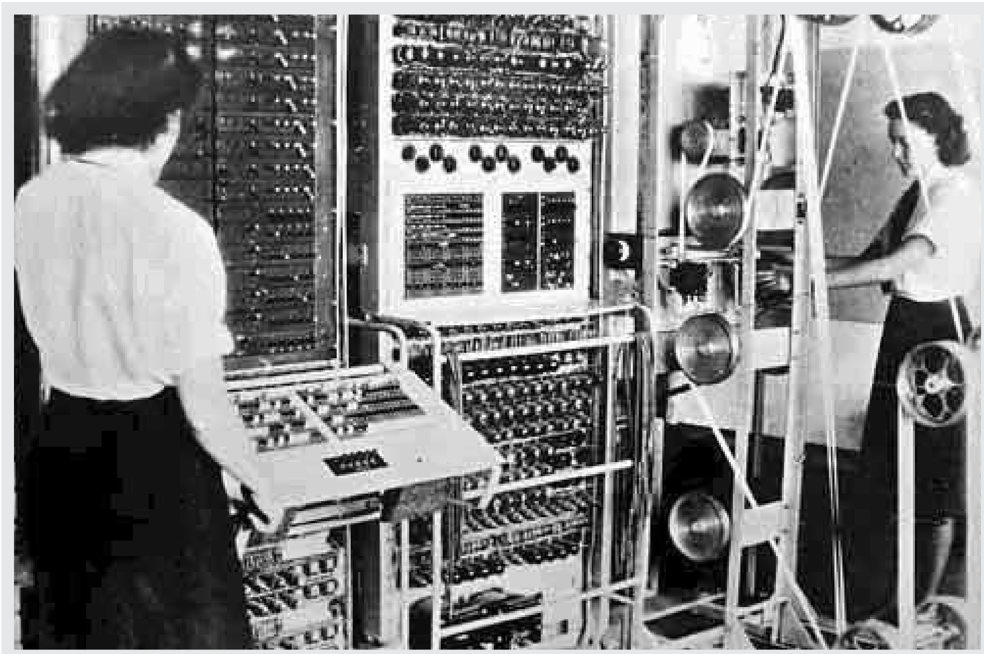


Figure 16.5: Colossus, the world's first totally electronic programmable computing device, built 1943-1945.

*Courtesy of The United Kingdom Government. Copyright: pd (Public Domain (information that is common property and contains no original authorship)).*

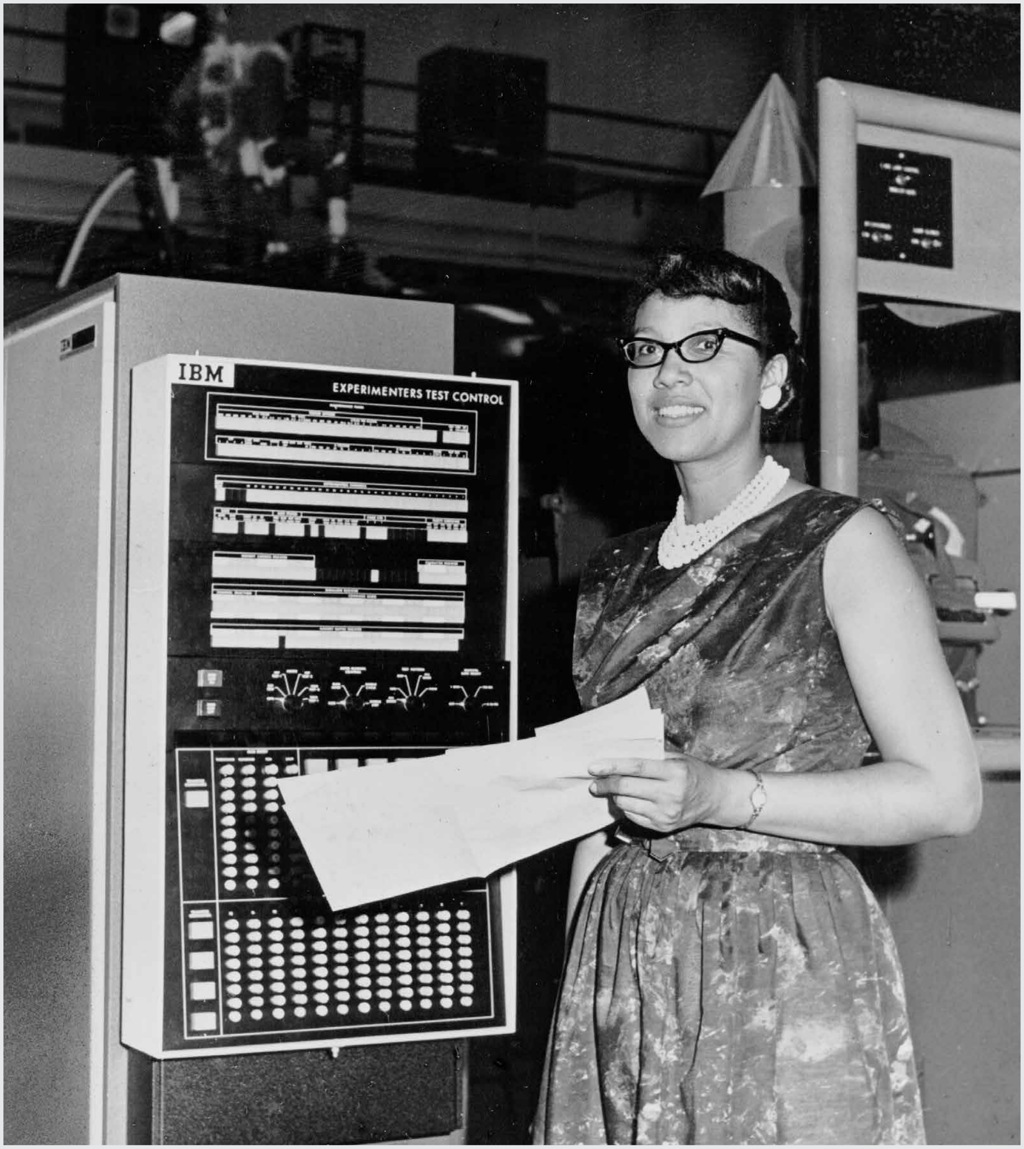



FIGURE 16.6: A female "computer": Melba Roy headed the group of NASA mathematicians who were known as "computers." They tracked the Echo satellites. Roy's computations helped produce the orbital element timetables by which millions could view the satellite from Earth as it passed overhead.

Figure 16.7: The Xerox Alto, introduced in 1973, but never commercially produced. The Alto was the predecessor of the Xerox Star, an early "personal computer", introduced in 1981.


FIGURE 16.8: A calculator on an iPhone, anno 2009.



### 16.3.2.5 Internalization and externalization

This principle states that human activities are distributed – and dynamically re-distributed – along the external/internal dimension. Any human activity contains both internal and external components. Sometimes external components are hardly visible: they can be reduced, for instance, to eye movements or even patterns of brain activation, but they are always present. The concepts of internalization and externalization refer to the processes of mutual transformations between internal and external components of an activity.

In the process of *internalization* external components become internal. For instance, young children often use their fingers to do simple math, but over time the use of fingers typically becomes redundant. An inexperienced driver may speak aloud to remind himself of the “parallel parking” procedure, but the need for speaking aloud is likely to disappear with practice.

The concept of internalization in activity theory is similar to the one proposed in some other frameworks, most notably Vygotsky’s cultural-historical psychology. Within Vygotsky’s framework the notion of internalization predominantly refers to a step in the development of higher mental functions, at which sign mediation initially emerging in the external plane eventually progresses to the internal plane. In activity theory internalization is used in a broader meaning as any re-distribution of internal and external components of an activity that results in a shift from the external to the internal.

Furthermore, in activity theory internalization is considered as just one type of transition, and providing a full account of the dynamics of activity “…necessarily presupposes the existence of regularly occurring transitions in the opposite direction also, from internal to external activity. “ (Leontiev, 1978). The process, opposite to internalization is *externalization* – that is, transformation of internal components of an activity into external ones. An example of externalization is sketching a design idea (see Figure 16.9).

Leontiev observed that in modern forms of work internal and external components of activity are becoming increasingly intertwined: ”Physical work accomplishing a practical transformation of material objects, ever more “intellectualized,” incorporates into itself the carrying out of more complex mental acts; at the same time the work of the contemporary researcher, activity that is specially cognitive, intellectual par excellence, is ever more filled with processes that in their form are external actions” (Leontiev, 1978).

In a similar vein, an activity, which is initially *socially* distributed, that is, distributed between several people (e.g., driving a car by a person taking a driving lesson, distributed between the learner and driving instructor) can be appropriated by a person (i.e., the learner) and then carried out individually. The opposite process is the transformation of an individual activity into a socially distributed one, e.g., when a person initiates a group project or other people intervene to help an individual to carry out her actions (Cole and Engeström, 1993). The dimensions of internal/external and individual/social are similar to one another in many respects and are closely related. For instance, when an internal activity is externalized, it also affects the individual-collective dimension: for instance, tools and signs employed in externally distributed actions can be shared and thus enable social distribution of the actions.

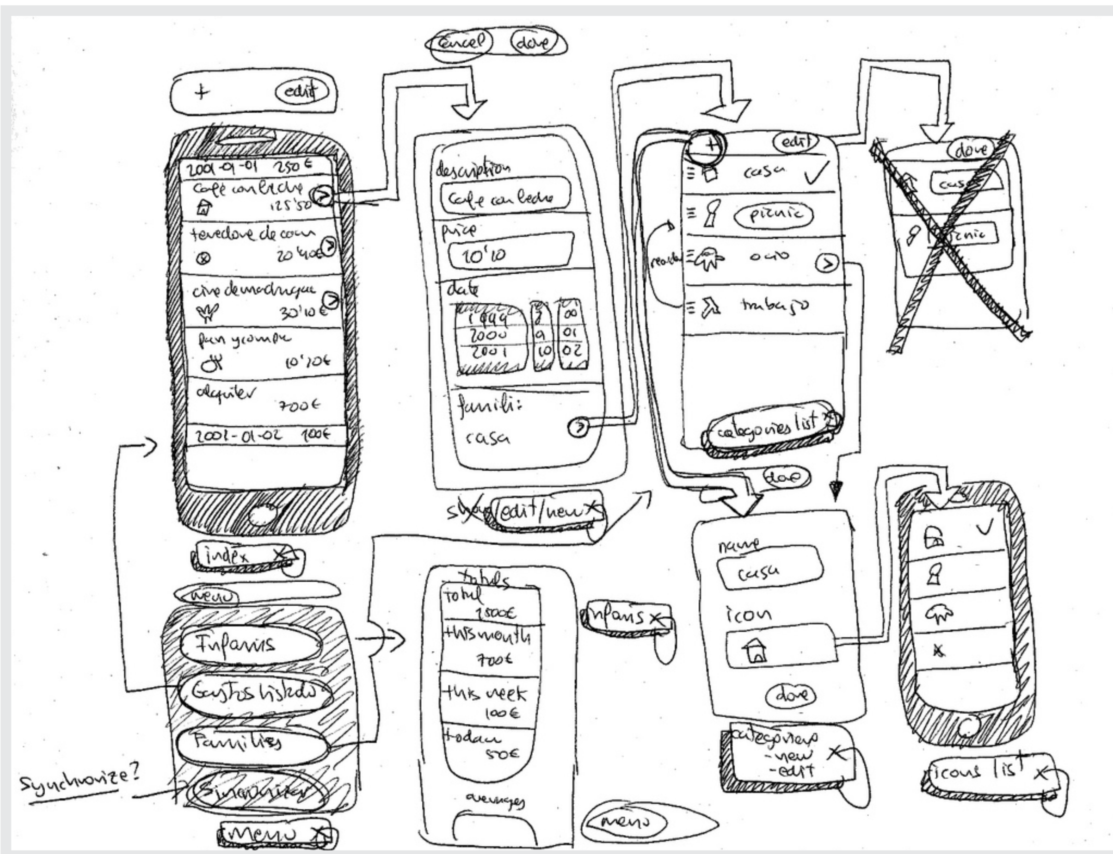



Figure 16.9: Externalization: Sketching a user interface idea.

### 16.3.2.6 Development

Finally, activity theory requires that activities always be analysed in the context of development. Development in activity theory is both an object of study and research strategy. As an object of study, development constitutes a complex phenomenon that can be analysed at different levels. Examples of the levels of analysis include studying the development of various forms of animal activity in biological evolution (phylogenesis), emergence of specifically human forms of activity in social history (sociogenesis), individual development throughout various phases of life (ontogenesis), appropriation of particular artefacts (instrumental genesis, Rabardel and Bourmaud, 2003), and so forth.

In activity theory development is also a research strategy. Analysis of the dynamics of how the object of study transforms over time is considered essential for a deep understanding of the object. Activity theory does not prescribe a single method of study since different types and levels of development require different methods or combinations of methods.

#### Dialectical logic

The developmental research perspective adopted by activity theory is often associated with *dialectical logic*, a concept and framework introduced by the Russian philosopher Evald Ilyenkov (Ilyenkov 2008; see Engeström et al., 1999). Dialectical logic is different from traditional formal logic in how it views contradictions and development. Traditional logic invariantly considers contradictions as indicators of problems that need to be addressed. Contradictions are to be eliminated in order to create a perfectly logical system (either an abstract one, such as a model or theory, or more concrete one, such as the management structure of an organization). In addition, traditional logic is typically not concerned with development; perfectly logical systems do not need to be changed and may stay as they are indefinitely.

Dialectical logic starts from a different assumption. It is assumed that dialectical development—that is, development driven by contradictions—is a fundamental aspect of all imaginable objects of study and therefore should be taken into consideration in analysis. While some "superficial" contradictions can be eliminated in a relatively straightforward way, there are also other, deeper contradictions which cannot be simply resolved once and for all. Any solution intended to resolve such contradictions is temporary, for it gives rise to new contradictions. An example of a contradiction of this type, well known to HCI researchers, is the contradiction between tasks and artefacts. The notion of "task-artefact cycle" (Carroll, 1991) implies that the ultimate balance between tasks and artefacts cannot be achieved. A new artefact changes the task for which it is developed which means that another artefact needs to be developed to support the new task, and so on and so forth.

Dialectical logic posits that analysis of the object of study which only deals with how the object exists at the present time is insufficient. Instead, analysis of the developmental trajectory of the object—preferably, starting from an initial undeveloped form (i.e., a "germ")—is claimed to be critically important for understanding how the object has come to be what it is, and what contradictions can be expected to drive its further development.

The principles of activity theory, described above, comprise an integrated system: they represent different aspects of human activity as a whole. Systematic application of any of the principles often makes it necessary to eventually engage the others as well. For instance, analysis of the effects of certain technologies on human cognition from an activity theoretical perspective would require identifying the variety of activities, as well as their respective objects within which the technologies are being employed (object-orientedness), the role and place of the technologies in the hierarchical structure of each of these activities (hierarchical structure), how the activities are being re-shaped by using the technologies as mediating means (mediation), and how transformations of external components of activity are related to corresponding changes of internal components (externalization and internalization). And all these phenomena should be analysed as they unfold over time (development).

### 16.3.3 Engeström's activity system model

Leontiev's approach is predominantly concerned with activities of individual human beings. While Leontiev explicitly mentions that activities can be carried out not only by individual human beings but also by social entities (collective subjects), too, he does not systematically explore the structure and development of collective activities and does not present a conceptual model of collective activity (which can probably be explained, at least partly, by the ideology-related limitations and constraints that were imposed on studies of social phenomena in the USSR). A model of collective activity, the "activity system model" (a.k.a. "Engeström's triangle") was proposed by the Finnish educational researcher Yrjö Engeström (1987). The model is a result of a two-step extension of Leontiev's original concept of activity— that is, activity understood as the "subject-object" interaction — to the case of collective activity.

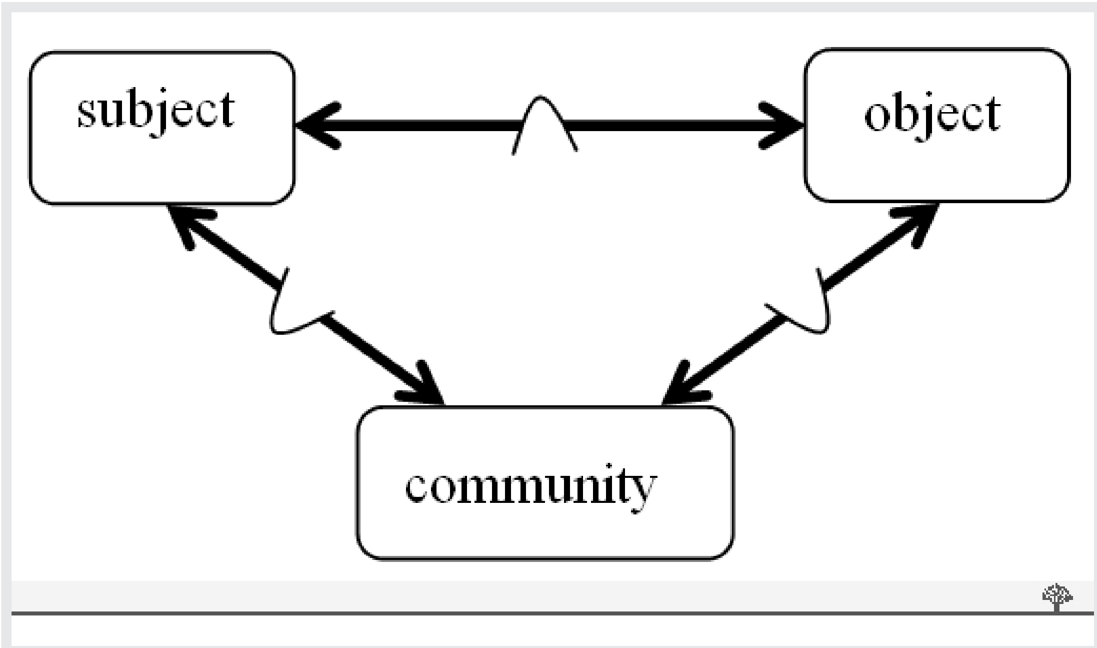



Figure 16.10: Three-way (mediated) interaction between subject, object, and community (adapted from Engestrom, 1987).



The first step, the most significant revision of Leontiev's notion of activity as the "subject-object" interaction, was adding a third node, "community", which resulted in a structure comprising a three-way interaction between "subject", "object", and "community". This structure can be represented as a down-pointing triangle (see Figure 16.10). Second, it was suggested that each of the three particular interactions within the structure is mediated by a special type of meditational means. Concrete mediational means for these interactions, according to Engeström, are: (a) *tools/ instruments* for the "subject - object" interaction (as also posited by Leontiev), (b) *rules* for the "subject – community" interaction, and (c) *division of labour* for the "community - object" interaction. In addition, the model includes the *outcome* of the activity system as a whole: a transformation of the object produced by the activity in question into an intended result, which can be utilized by other activity systems. The complete model is shown in Figure 16.11.

As an example, consider the activity of an interaction designer who works as a member of a design team on redesigning the user interface of a computer application. The object of the activity is the existing interface, and the expected outcome is a new interface. The interaction designer employs a variety of *tools* in her work on the object, including physical objects (e.g., computers), software (e.g., development environments), and methods and techniques (e.g., personas). The community comprises other members of the team: interaction designers, the project manager, technicians, etc. The interaction designer's relation with the community is mediated by explicit and implicit rules, e.g., taking part in project meetings, receiving certain financial rewards, etc. Furthermore, producing the outcome of the activity system as a whole, a new interface, is the responsibility of the entire design team: the effort of the interaction designer is a part of a larger effort of the team. Therefore, the work of the interaction designer needs to be coordinated with the work of other team members. This coordination is achieved by employing a division of labour, which thus mediates the relation between the design team and its object.

When studying complex real-life phenomena, applying one activity system model is often not sufficient. Such phenomena need to be represented as *networks of activity systems*. For instance, redesigning the user interface of a computer application can be a part of an even larger-scale effort, involving several design teams, directed at developing a new version of the computer application in question. Redesigning the user interface in that case would provide a partial outcome which would need to be integrated with outcomes of other activity systems (e.g., a team developing new functionality of the product) to achieve the overarching purpose of a network of activity systems.

A key tenet of Engeström's framework is that activity systems are constantly developing. The development is understood in a dialectical sense as a process driven by contradictions. Engeström identifies four types of contradictions in activity systems:

1. Primary contradictions are inner contradictions of each of the nodes of an activity system. For instance, the mediating means used by a physician include various medications which, on the one hand, have certain medical effects, and, on the other hand, are products with associated costs, legal regulations, distribution channels, etc.

2. Secondary contradictions are those that arise between the nodes of an activity system. For instance, a certain type of medical treatment may be unsuitable for certain patients

3. Tertiary contradictions describe potential problems emerging in the relationship between the existing forms of an activity system and its potential, more advanced object and outcome. The advancement of an activity system as a whole may be undermined by the resistance to change, demonstrated by the existing organization of the activity system.

4. Finally, quaternary contradictions refer to contradictions within a network of activity systems, that is, between an activity system and other activity systems involved in the production of a joint outcome.

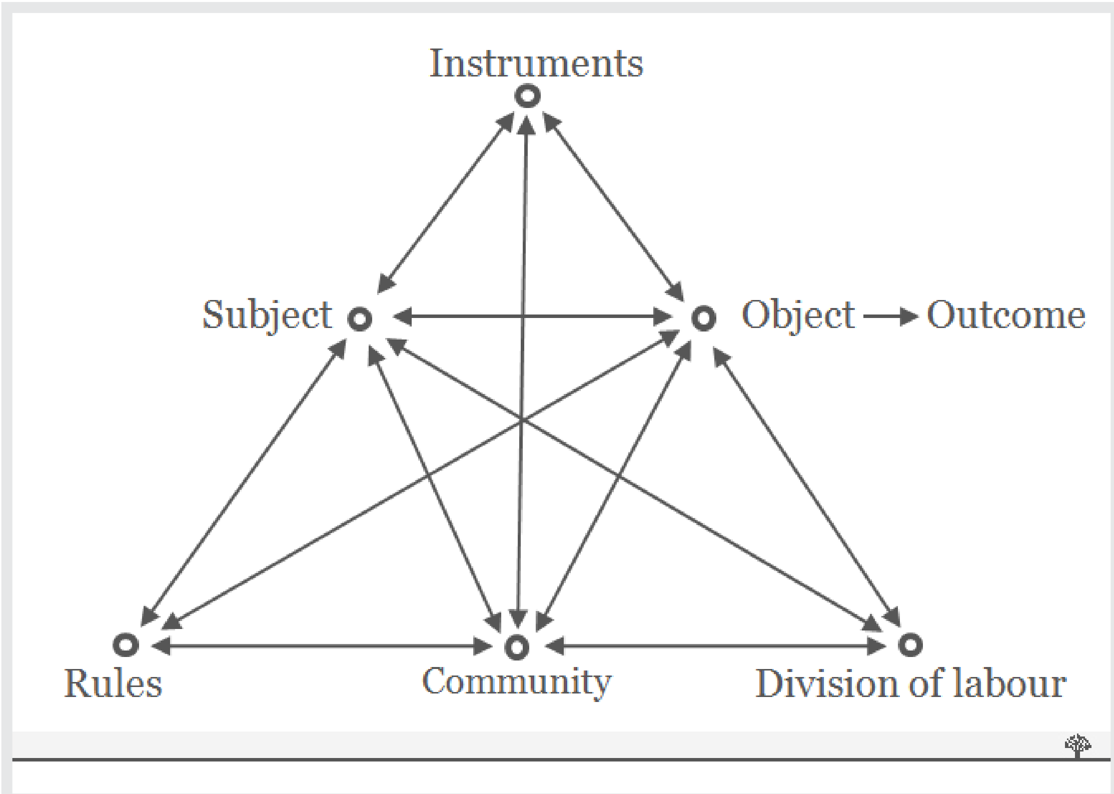



Figure 16.11: Engestrom's activity system model.



The activity system model has been employed in a range of disciplines, especially education and organizational learning (see, e.g., CRADLE, 2011).

### 16.3.4 Current diversity of activity theoretical frameworks

The approaches developed by, respectively, Leontiev and Engeström are currently the most common variants of activity theory. The approaches provide complementary perspectives on human activities. Leontiev's variant mostly focuses on individuals understood as social creatures acting in social contexts. Engeström's activity system model, on the other hand, is predominantly concerned with collective activities carried out by groups and organizations and implemented through contributions—i.e., actions—of individual subjects.

In addition, a number of other current frameworks are partly influenced by activity theory and partly built upon other approaches. Such frameworks include, for instance, instrumental genesis (Rabardel and Bourmaud, 2003), genre tracing (Spinuzzi, 2003), and the systemic-structural activity theory (Bedny and Harris, 2005; Bedny and Karwowski, 2003).

## 16.4 ACTIVITY THEORY IN HCI AND INTERACTION DESIGN

### 16.4.1 Activity theory as a second-wave, post-cognitivist HCI theory

The dominant paradigm in HCI when it appeared as a field in early 80s was information processing ("cognitivist") psychology. But the HCI community gradually came to realize that the focus on information processing was not sufficient. Individuals' interests, needs, frustrations, and so forth, proved to be important and powerful factors in choosing, learning, and using a technology. Furthermore, it was becoming increasingly obvious that the use of technology critically depends on complex, meaningful, social, and dynamic contexts in which it takes place. The inner logic of the development of the field required that the scope of HCI be expanded to include the issues of motivation, meanings, culture, and social interactions. However, the cognitivist approach could not provide conceptual tools for dealing with such issues. When the limitations of the information processing psychology in HCI became widely acknowledged (Carroll, 1991), activity theory was identified as a potential alternative theoretical foundation for the field (Bødker, 1991).

The impact of activity theory on HCI and interaction design in the last two decades has been, essentially, threefold. First, the theory offered some general theoretical insights that resonated with the need for a richer conceptual framework which would allow the field to move from the "first-wave HCI" to the "second-wave HCI" (see Cooper and Bowers, 1995). Second, it served as an analytical framework for design and evaluation of concrete interactive systems and stimulated the development of a variety of analytical tools. Third and finally, the application of the approach, especially in recent years, resulted in a number of novel systems, implementing the ideas of activity-centric (or activity-based) computing.

### 16.4.2 General theoretical insights

The unit of analysis proposed by activity theory, that is, "subject-object" interaction, may appear similar to the traditional focus of HCI on "human-computer" interaction. However, adopting an activity theoretical perspective had important implications for understanding how people use interactive technologies. First of all, it made it immediately obvious that "computer" is typically not an *object* of activity but rather a mediating artefact. Therefore, generally speaking, people are not interacting *with* computers: they interact with the world *through* computers. The book by Susanne Bødker, which played a key role in introducing activity theory to HCI, reflected this perspective on interactive technologies in its very title: "Through the interface: An activity-theoretical perspective on human-computer interaction" (Bødker, 1991).

Another general theoretical contribution of activity theory to HCI was placing computer use in the hierarchical structure of human activity, that is, relating the operational aspects of the interaction with technology to meaningful goals and, ultimately, needs and motives of technology users. It did not mean rejecting the formal models of users and tasks which were developed in early HCI research, but rather extending the scope of analysis beyond low-level interaction. Such an extension was considered by some researchers as perfectly consistent with the need of the field to move "from human factors to human actors" (Bannon, 1991).

Finally, adopting the conceptual framework of activity theory promised to open up new possibilities for analysing the *context* of technology use. As mentioned, the lack of conceptual tools for understanding context was a major limitation of the information-processing psychology in HCI. Activity theory, with its emphasis on society, culture, and development, offered a set of concepts for capturing the context of use and taking it into account in the design, evaluation, and deployment of interactive technologies. An edited collection entitled "Context and consciousness: Activity theory and human-computer interaction" (Nardi, 1996a) provided an in-depth exposition of a wide range of such concepts.

### 16.4.3 Activity theory and other 'second-wave' theories

There are both similarities and differences between activity theory and other "second-wave theories" (cf. Kaptelinin et al., 2003), such as phenomenology (Winograd and Flores, 1986; Svanaes, 2000; Dourish, 2001) and distributed cognition (Hollan et al, 2000; Rogers, 2004).

A fundamental assumption uniting most second-wave theories is that human beings cannot be understood separately from the world in which they live, act, and cognize. The need to analyse the inseparability of humans and their physical, social, and information environments is emphasized by activity theory's notion of activity as "subject-object" interaction, phenomenology's concept of "being-in-the-world" (Dourish, 2001), and distributed cognition's models of the propagation of representations across the boundaries of humans and artefacts (Hollan et al., 2000; Rogers, 2004). Another key notion, common to many post-cognitivist frameworks, is that humans' connection to the world is to a large degree determined by the artefacts used by the humans, which artefacts are variously defined in terms of mediating means (activity theory), equipment (phenomenology), or processing and transmission of information (distributed cognition and external cognition) (see also Nardi, 1996a; Kaptelinin and Nardi, 2006).

While similar in a number of important respects, "second-wave" theories are also different in their general perspectives on humans and human relation to the world. Phenomenology is relatively less interested in the issues of development and the social nature of human beings, compared to activity theory. For instance, the question of how "being-in-the world" comes to exist in the first place (that is, how exactly we are thrown into the world) does not seem to play a critical role in phenomenology, which is in stark contrast with the attention paid in activity theory to how subjects emerge in evolution, social history, and individual development. In addition, a systematic exploration of the social dimension of being is a relatively recent development in the phenomenological tradition (Dourish, 2001), even though the need to take it into account was already emphasized, for instance, in the foundational work of Heidegger (1962).

The distributed cognition framework, at least as it is applied in HCI, is less explicit in its general assumptions about the nature of human beings and mostly focuses on concrete problems of understanding and supporting cognitive processes distributed between people and artefacts (Rogers, 2004). It is, however, apparent that activity theory and distributed cognition substantially differ in their respective views of human agency. Human agency is a major conceptual point of departure in activity theory, while in HCI research informed by the distributed cognition framework this issue does not play a significant role.

Comprehensive analysis of similarities and differences between activity theory and other post-cognitivist approaches is a complex issue, which is beyond the scope of this chapter. Such analysis can be found elsewhere. A systematic comparison of activity theory with a variety of other approaches is conducted by Nardi (1996b) and Kaptelinin and Nardi (2006). Halverson (2002) discusses activity theory and distributed cognition as theoretical frameworks for CSCW research. Rogers (2004) provides an overview of current theoretical approaches in HCI, including activity theory, distributed cognition, and external cognition.

### 16.4.4 Analytical tools

Activity theory is not a "theory" in the traditional sense in which "theory" is understood in natural sciences. Activity theory does not support creating and running predictive models which would only need to be "fed" with appropriate data. Instead, it aims to help researchers and practitioners to orientate themselves in complex real-life problems, identify key issues which need to be dealt with, and direct the search for relevant evidence and suitable solutions. In other words, the key advantage of activity theory appears to be in supporting researchers and practitioners in their own inquiry—for instance, by helping to ask right questions—rather than providing ready-made answers.

A variety of analytical tools, informed by activity theory, have been proposed to support asking "the right questions" in analysis, design, and evaluation of interactive systems (Quek and Shah, 2004). Most of such tools have the format of a checklist: they are, essentially, organized lists of questions or issues that researchers or practitioners need to pay attention to in order to make sure that the most important aspects of human activity are taken into account. The choice of the checklist format is intended to help bridge the gap between theory's high level of abstraction and the need to address concrete issues in analysis and design. Arguably, the elaborated system of concepts (and their relations) offered by activity theory can be used in HCI to better understand the role and place of concrete interactive technologies in the overall structure of purposeful, mediated, social human action. However, the framework provides a high level description, not limited to particular types of artefacts, and needs to be specifically adjusted to the requirements of HCI research and practice. Such an adjustment can, in principle, be delegated to HCI researchers and practitioners themselves, but in many cases this strategy may not be realistic since it would require considerable time and effort. Activity theory-based checklists reduce the effort associated with domain-specific adjustment of the theory by converting the organized set of concepts, offered by the theory, into a set of concrete issues and questions, specifically related to analysis and design of interactive technologies.

Different types of such checklists are based on different variants of activity theory. For instance, the Activity Checklist (Kaptelinin et al., 1999) is intended to support systematic exploration of the "space of context" in design and evaluation of interactive technologies. The overall structure of the checklist is derived from the basic principles of Leontiev's framework. The checklist comprises four sections—Means and ends, The environment, Learning, cognition and articulation, and Development—which are produced by combining the principle of *mediation* with, respectively, the principles of object-orientedness, the hierarchical structure of activity, internalization/externalization, and development, The checklist was employed in a number of design and evaluation projects (see Kaptelinin and Nardi, 2006).

Jonassen and Rohrer-Murphy (1999) introduce another analytical tool, based on a somewhat different (while partly overlapping) set of activity-theoretical concepts. The tool comprises several organized arrays of questions and issues mostly derived from Engeström's activity system model. The basic components of the model—Subject, Object, and Community, as well as Tools, Rules, and Roles mediating the three-way interaction between the components—serve as the main rubric for issues that need to be taken into account and modeled when designing the components of a constructivist learning environment, as well as the relationship between the components. The AODM (Activity-oriented design method) approach to supporting technology-enhanced learning analysis and design, developed by Mwanza (2002) includes several lists of issues to explore, which lists mostly capitalize upon the conceptual structure provided by Engeström's activity system model. For instance, the Eight-Step Model prescribes a sequence of analytical steps, starting from focusing on the activity system in question as a whole, then proceeding to each of the six individual nodes and, finally, analysing the outcome of the activity system.

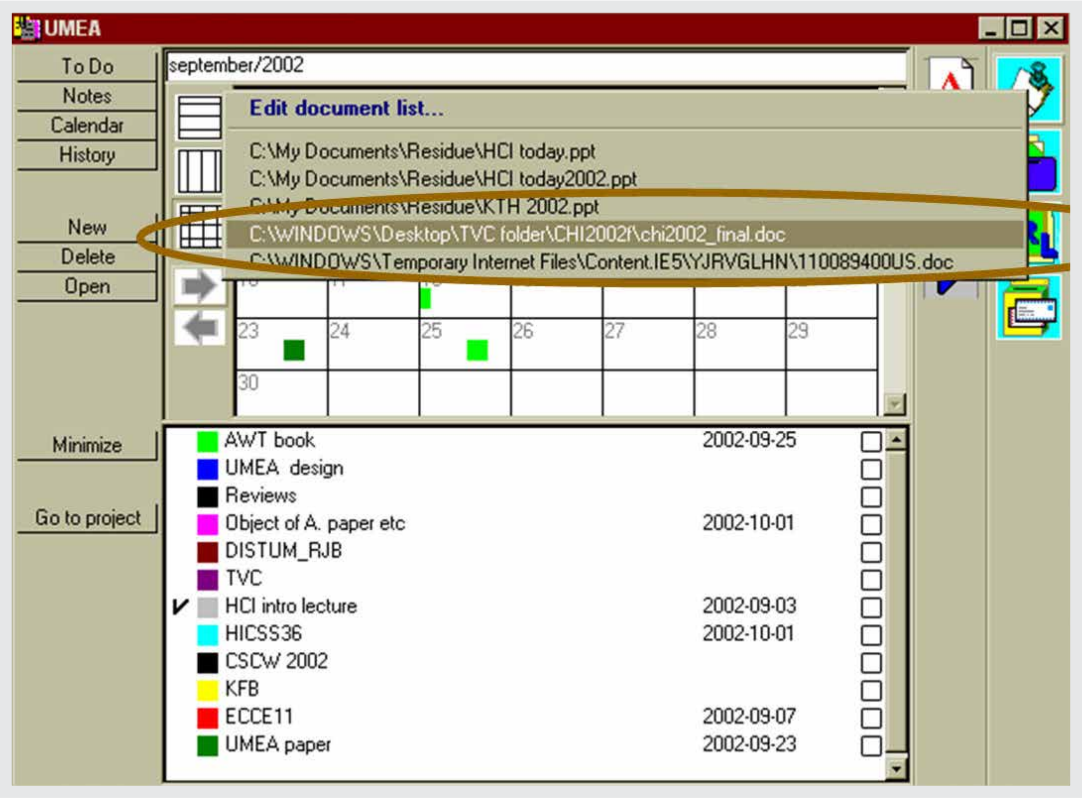



FIGURE 16.12: User interface of the UMEA system (Kaptelinin, 2003).

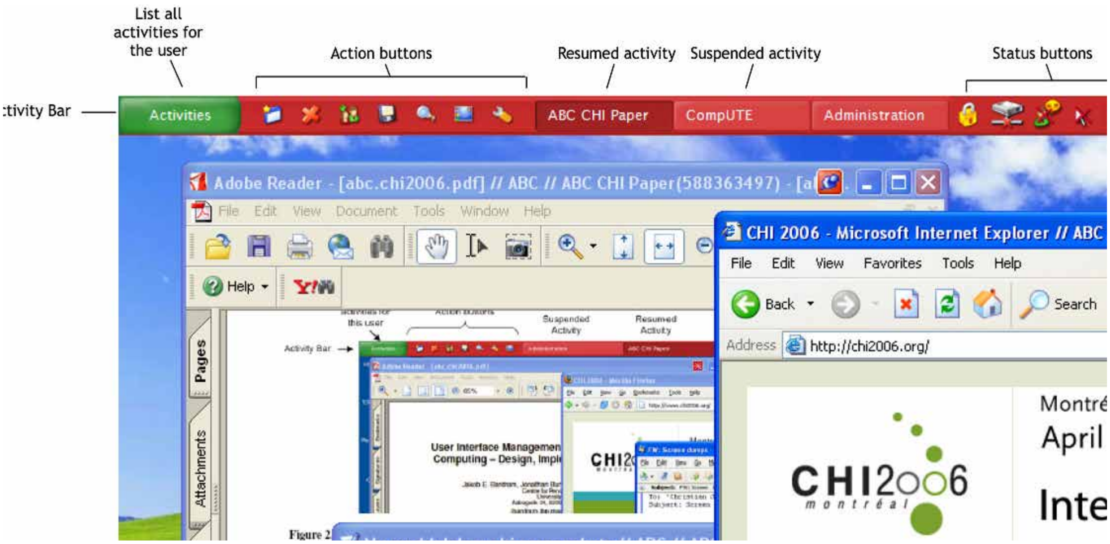



Figure 16.13: An overall view of the ABC user interface for Windows XP (Bardram et al., 2006).

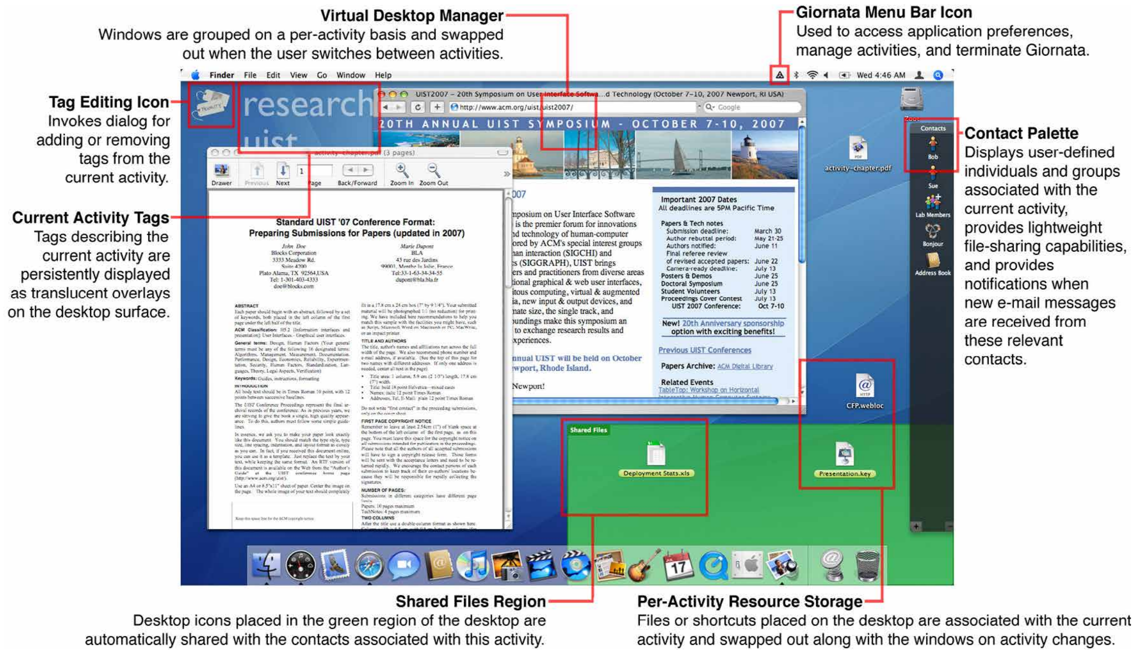



FIGURE 16.14: Giornata's user interface.

*Courtesy of Stephen Voida. Copyright: CC-Att-ND (Creative Commons Attribution-NoDerivs 3.0 Unported).*

## 16.4.5 Activity-centric computing

Adopting an activity-theoretical perspective has an immediate implication for design: it suggests that the primary concern of designers of interactive systems should be supporting meaningful human activities in everyday contexts, rather than striving for logical consistency and technological sophistication. Currently many systems fail to comply with this, seemingly obvious, requirement. For instance, traditional desktop systems organize digital resources into formal categories (e.g., files, email messages, bookmarks...) rather than according to the relevance of a resource to the task at hand, and most systems provide limited support for task switching and interruptions (Bardram et al., 2006; Kaptelinin and Czerwinski, 2007).

Activity-centric (also referred to as "activity-centred" or "activity-based") computing is an approach to designing interactive systems according to which the top priority and an explicit aim in the design of digital artefacts and environments should be supporting meaningful human activities. The work in activity-centric computing is being conducted from a diversity of perspectives; some of the key projects (e.g., Moran et al., 2005; Moran, 2006) do not employ actvity theory as their theoretical foundation. It is fair to say, however, that the theory has influenced, one way or another, many (if not most) developments in the area.

An early attempt to propose an activity-centric alternative to then dominant application-centric and document-centric approaches was made by Don Norman and his colleagues at Apple Computer (Norman, 1998), who employed a somewhat modified version of Leontiev's framework. More recently, Norman (2005) argues that activity-centred design has advantages over traditional human-centred design and should supersede the latter.

For various reasons, the attempt to introduce an activity-centric approach at Apple Computer has not resulted in the development of concrete novel technologies. However, in the recent decade a number of systems, adopting an activity-centric perspective and, for the most part, explicitly informed by activity theory, have been designed and implemented. They include, for instance, the UMEA system (Kaptelinin, 2003, see Figure 16.12), a variety of systems implementing the ABC framework, e. g. the Windows XP ABC system (Bardram et al., 2006, see Figure 16.13), and the Giornata system (Voida and Mynatt, 2009, see Figure 16.14). All these systems provide alternatives to, or extensions of, traditional desktop systems to enable organizing various digital resources, such as documents and URL's, around higher-level, meaningful tasks of the user, defined as "projects" or "activities". In UMEA and the Windows XP ABC system it is achieved by automatically assigning resources to the activity selected by the user, while in Giornata a virtual desktop is set up for each new activity.

The results of evaluation studies (Kaptelinin, 2003; Bardram et al., 2006; Voida and Mynatt, 2009) suggest that activity-centric systems have certain advantages over more conventional types of systems.

| Search string | Number of hits | Search string | Number of hits |
|---|---|---|---|
| phenomenology | 1,552 | phenomenology & HCI | 251 |
| “activity theory” | 1,496 | ”activity theory” & HCI | 512 |
| “distributed cognition” | 1,102 | “distributed cognition” & HCI | 383 |
| ethnomethodology | 621 | ethnomethodology & HCI | 269 |
| ”situated action” | 551 | ”situated action” & HCI | 209 |
| “language action” | 448 | “language action” & HCI | 66 |
| “actor network theory” | 374 | “actor network theory” & HCI | 43 |
| ”external cognition” | 141 | “external cognition” & HCI | 60 |

TABLE 16.1: Second wave theories: Number of hits in the ACM Digital Library (dl.acm.org) for names of a selection of theoretical approaches used as search strings, January 2nd, 2012.

## 16.5 CONCLUSIONS AND PROSPECTS FOR THE FUTURE

The discussion in this chapter indicates that in the last two decades, since its introduction to HCI, activity theory has established itself as a leading theoretical approach in the field. Along with some other post-cognitivist approaches, most notably distributed cognition (Hollan et al., 2000) and phenomenology (Svanaes, 2000; Dourish, 2001), it shapes the theoretical landscape of current HCI and interaction design. A number of fundamental notions, such as technological mediation, originating from activity theory have become widely accepted in the field. Table 16.1 gives an approximation, if very rough and imprecise, of the relative “popularity” of activity theory in studies of information technologies compared to some other theoretical approaches.

At the same time, lessons learnt from applying activity theory in HCI and interaction design indicate that the theory needs to be further developed, and there are some issues which must be addressed in the future. First, the concepts of activity theory should be more clearly specified and operationalized to make it easier for researchers and practitioners to see how the theory can be applied in concrete cases (cf. Rogers, 2004). Second, the conceptual framework of activity theory needs to be expanded to more adequately deal with coordination of multiple activities and cross-activity integration. Cross-activity integration is becoming an increasingly important issue in current uses of technology, characterized by complex social contexts (e.g., a combination of work and non-work factors typical of everyday practices of teleworkers) and employing multiple digital and non-digital technologies (or “webs of mediators” – see Bødker and Andersen, 2005). Third, as observed by Bødker (2006), HCI appears to be entering its new, third wave, during which there is a marked increase of interest in aesthetics and experience (see also Hassenzahl, 2011). The move toward third-wave HCI, according to Bødker (2006) presents a challenge for second-wave theories, including activity theory. Arguably, the conceptual apparatus of activity theory can, in principle, be employed to analyse subjective experiences, and some activity-theoretical analyses do address the issues of emotion, passion, and so forth (e.g., Vasilyuk, 1992; Nardi, 2005). However, the potential of activity theory to deal with such issues remains relatively untapped. Expanding the scope of activity-theoretical analysis in these three directions appears to be essential to make sure the theory continues to provide HCI with new insights and to help the field to deal with emerging challenges.

# 16.6 WHERE TO LEARN MORE

## 16.6.1 Activity theory in HCI

### 16.6.1.1 Books


Bødker, Susanne (1991): *Through the Interface - A Human Activity Approach to User Interface Design*. Hillsdale, NJ, Lawrence Erlbaum Associates

Gay, Geraldine and Hembrooke, Helene (2004): *Activity-Centered Design: An Ecological Approach to Designing Smart Tools and Usable Systems (Acting with Technology)*. The MIT Press

Kaptelinin, Victor and Nardi, Bonnie A. (2012): *Activity Theory in HCI: Fundamentals and Reflections*. Morgan and Claypool

Kaptelinin, Victor and Nardi, Bonnie A. (2006): *Acting with Technology: Activity Theory and Interaction Design*. The MIT Press

Nardi, Bonnie A. (ed.) (1996): *Context and Consciousness: Activity Theory and Human-Computer Interaction*. MIT Press

Spinuzzi, Clay (2003): *Tracing Genres through Organizations: A Sociocultural Approach to Information Design (Acting with Technology)*. The MIT Press


### 16.6.1.2 ARTICLES AND BOOK CHAPTERS


Baerentsen, Klaus B. and Trettvik, Johan (2002): An activity theory approach to affordance. In: Proceedings of the Second Nordic Conference on Human-Computer Interaction October 19-23, 2002, Aarhus, Denmark. pp. 51-60

Beaudouin-Lafon, Michel (2000): Instrumental Interaction: An Interaction Model for Designing Post-WIMP User Interfaces. In: Turner, Thea, Szwillus, Gerd, Czerwinski, Mary, Peterno, Fabio and Pemberton, Steven (eds.) Proceedings of the ACM CHI 2000 Human Factors in Computing Systems Conference April 1-6, 2000, The Hague, The Netherlands. pp. 446-453

Bedny, Gregory and Karwowski, Waldemar (2003): A Systemic-Structural Activity Approach to the Design of Human-Computer Interaction Tasks. In International Journal of Human-Computer Interaction, 16 (2) pp. 235-260

*Bertelsen, Olav W. and Bødker, Susanne (2003): Activity Theory. In: Carroll, John M. (ed.). "HCI Models, Theories, and Frameworks". San Francisco: Morgan Kaufman Publisherspp. 291-324*

*Blackler, F. (1995): Activity theory, CSCW and organizations. In: Monk, Andrew F. and Gilbert, Nigel G. (eds.). "Perspectives on HCI: Diverse Approaches". Academic Press*

*Kuuti, Kari (1995): Activity theory as a potential framework for human-computer interaction research. In: Nardi, Bonnie A. (ed.). "Context and Consciousness: Activity Theory and Human-Computer Interaction". The MIT Presspp. 17-44*

*Rogers, Yvonne (2004): New Theoretical Approaches for HCI. In Annual Review of Information Science and Technology, (38) pp. 1-43*

*Wilson, T. D. (2008): Activity theory and information seeking. In Annual Review of Information Science and Technology, 42 (1) pp. 119-161*

## 16.6.2 Activity theory in general

### 16.6.2.1 Books

*Cole, Michael (1998): Cultural Psychology: A Once and Future Discipline. Cambridge, MA, The Belknap Press of Harvard University Press*

*Daniels, Harry, Edwards, Anne, Engeström, Yrjö, Gallagher, Tony and Ludvigsen, Sten R. (eds.) (2009): Activity Theory in Practice: Promoting Learning Across Boundaries and Agencies. Routledge*

*Engeström, Yrjö (2008): From Teams to Knots: Activity-Theoretical Studies of Collaboration and Learning at Work (Learning in Doing: Social, Cognitive and Computational Perspectives). Cambridge University Press*

*Engeström, Yrjö (ed.) (1990): Learning, Working and Imagining. Orienta-Konsultit Oy*

*Engeström, Yrjö (1987): Learning by expanding: An activity-theoretical approach to developmental research. Orienta-Konsultit Oy*

*Engeström, Yrjö, Miettinen, Reijo and Punamaki, Raij (eds.) (1999): Perspectives on Activity Theory (Learning in Doing Social, Cognitive and Computational Perspectives). Cambridge University Press*

*Koschmann, Timothy, Kuutti, Kari and Hickma, Larry (1998): The Concept of Breakdown in Heidegger, Leont'ev, and Dewey and Its Implications for Education. In Mind, Culture, and Activity, 5 (1) pp. 25-41*

*Leontiev, Aleksei Nikolaevich (1978): Activity, consciousness, and personality. Prentice-Hall*

*Leontiev, Aleksei Nikolaevich (1981): Problems of the development of the mind (originally published in Russian in 1959). Moscow, Russia, Progress*

*Rubinshtein, Sergey L. (1946): Foundations of General Psychology. Second edition. (in Russian). Moscow, Russia, Uchpedgiz*

*Vasilyuk, Fyodor (1992): The Psychology of Experiencing: The Resolution of Life's Critical Situations (originally published in Russian in 1984). New York University Press*

*Wertsch, James V. (1981): The Concept of Activity in Soviet Psychology. M.E. Sharpe*

### 16.6.2.2 ONLINE RESOURCES

- The Mind, Culture, and Activity Homepage is an interactive forum for a community of interdisciplinary scholars who share an interest in the study of human mind in its cultural and historical contexts.
- International Society for Cultural and Activity Research (ISCAR)

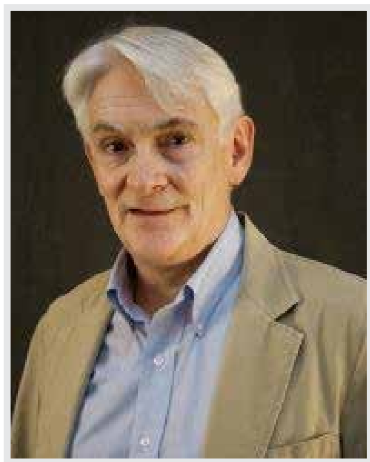
## 16.7 ACKNOWLEDEGEMENTS

I would like to thank Mads Soegaard, Rikke Friis Dam, Klaus Bærentsen, Allan Holstein-Rathlou, Bonnie Nardi, and Dag Svanæs for their feedback, help, and support. Special thanks to Mads for providing numerous figures for this chapter

## 16.8 COMMENTARY BY JOHN M. CARROLL

How to cite this commentary in your report

### John M. Carroll



© *John M. Carroll*

John M. Carroll is Edward Frymoyer Chair Professor of Information Sciences and Technology at the Pennsylvania State University. His research interests include methods and theory in human-computer interaction, particularly as applied to networking tools for collaborative learning and problem solving, and the design of interactive information systems. His books include Making Use (MIT P...

John M. Carroll
John M. Carroll is a member of The Interaction Design Foundation

### 16.8.1 Activity Awareness

Activity-centric computing is exploring the idea of activity as an appropriate organizing rubric for user interfaces, in contrast to more traditional user interface designs organized around applications and data type hierarchies. This is an important direction for user interface design, and a specific design implication drawing upon Activity Theory. In my work, I am exploring the idea that collaborators need to share awareness of joint activities – in contrast, for example, to conceiving of awareness only with respect to joint actions, mutual presence, and/or shared synchronous situations.

Collaborators must attain and maintain reciprocal *awareness* in order to coordinate effectively (Dourish & Bellotti, 1992). Groups engaged in collaborative activities of significant scope and duration must achieve and maintain *awareness* of diverse aspects of their shared activity in order to coordinate effectively. For example, they must verify mutual presence and attention, which is fairly straightforward in face-to-face interaction, but often subtle, difficult, and a continuing challenge in computer-mediated collaboration. Members need to know what tools and resources they have access to, but also what tools and resources their counterparts can access. The availability of tools and resources may change throughout the course of an activity. The group must have an understanding of who among them might know potentially relevant information, or know how to do something that might be critical to the collective endeavor. Members need to know something of their partners' attitudes and goals, and of what their partners expect from them and of the activity. They need to know what criteria their partners will use to evaluate joint outcomes, the moment-to-moment focus of their attention and action during the collaborative work, and how the view of the shared plan and the work actually accomplished evolves over time. All of these intentional variables change constantly as the task context itself changes.

Awareness in collaborative situations is sometimes regarded as a relatively discrete achievement – awareness of a task context (situation awareness), of group consensus, or of a shared mental model. These simplifications can be useful for scripted collaborative tasks, such as managing single-threaded processes or team training exercises. However, they do not address routine sources of complexity. In realistically complex tasks of significant scope and duration, the *current situation* is defined to a considerable extent by its history, which in turn is constantly reconstructed by the group and by its individual members. For instance knowing how other group members respond to criticisms can have a profound effect on group discussion and argumentation. The current situation is defined also by continuous exogenous dynamics that present a constantly changing situation to the group. Indeed, if awareness were to be supported by discrete updates, it would require an unceasing torrent of information, which ipso facto could never be useful or even usable.

Shared mental models are a popular way to think about the knowledge and skills that teams use to manage collective activity. But the notion of identical copies of knowledge used and maintained by team members to enable coordination is both exotic and cumbersome as a foundation for joint endeavor. Team members who believe that they should hold exactly the same understanding of a current task might spend considerable time and effort verifying agreed-upon preconditions for action, making them less useful to their partners in action than members who have different perspectives, and who could play complementary roles and take complementary team responsibilities. Moreover, too much literal shared understanding could entrain redundant capabilities, and teams no better than their best member. Teams with homogeneous understandings are maximally vulnerable to groupthink and stagnant thinking. Analogous to arguments regarding natural selection, the more variation that exists in a team, in individual backgrounds, mind sets and strategic approaches, the better the chances of that team to adapt to new and novel situations. For realistic and complex one-of-kind situations, such as emergency response, information analysis, and software design, creativity, learning and adaptation are critical to team performance. We are trying to articulate a sense of shared understanding among team members that is robust with respect to exogenous dynamics, and that can, in principle, leverage collaboration to produce performance better than any team member.

My colleagues and I are developing the concept of *activity awareness* as a programmatic analysis for the mutual awareness of partners sharing an activity of significant scope and duration (Carroll et al. 2003, 2006, 2009, 2011). Activity awareness builds upon, but transcends, synchronous awareness of where a partner's cursor is pointing, where the partner is looking, and other immediate features of a task situation. More importantly, it transcends the sharing of identical states of situation awareness or mental models. Indeed, we would argue that lower-level and simpler aspects of awareness are appropriately conceptualized as mediated by shared mental models: All stakeholders in a joint activity must have the same understanding of primitive and objective situation properties such as the document being edited, the key that was pressed, the reference of a deictic. But shared mental models are neither useful nor possible for intentional situation properties such as role-based interpretations and strategies, personal insights and perspectives, opportunistic problem solving derived from interactions with tools and other resources, value-based assessments drawing on personal histories, expectations and attributions about one's teammates, etc.

In framing activity awareness, we appropriated the concept of *activity* from Activity Theory, to emphasize that collaborators need be aware of a whole, shared activity as complex, socially and culturally embedded endeavor, organized in dynamic hierarchies, and not merely aware of the synchronous and easily noticeable aspects of the activity. In this view, awareness is teleologically inseparable from collective regulation of a joint endeavor. Members need to be engaged with one another's interests, values, and possibly relevant knowledge and skills, initial and current goals and motivations, criteria for evaluating outcomes, and assessments of the status and trajectory of ongoing work. This engagement is continually negotiated and developed. We articulated this continual process of activity awareness into arenas of conceptual negotiation among members of a team, a collection of ongoing interaction protocols rather than static sources of knowledge. Ours is a developmental framework in the traditional sense of Piaget and Vygotsky: higher-level facets are enabled by and resolve conflicts in lower-level facets.

When people plan, negotiate and coordinate with others in open-ended endeavors over significant spans of time, when they solve problems that are ill defined and consequential, when they stretch their own capabilities, they develop; that is, they come to experience and interact with the world in new ways. In Activity Theory, *human development* is a normal outcome of significant activity, but it is also profound in the sense that it qualitatively changes one's awareness of activity. As an individual develops, he or she becomes more able to understand, to reconcile, and to integrate different levels of performance and different approaches to problems by synthesizing zones of proximal development. The successive elaboration of personal perspectives further enhances each member's awareness of his or her own activity, and creates myriad new ways to construct common ground, codify practices, and build social capital. A shorthand for activity awareness is a group's awareness and regulation of its own activity.

Activity awareness is fundamentally a dynamic process, not a state of knowledge. It involves monitoring and integrating many different kinds of information at different levels of analysis, such as events, tasks, goals, social interactions and their meanings, group values and norms, and more. It involves monitoring and integrating more-or-less continuingly to learn about developing circumstances and the initiatives, reactions, and sense making of other people with respect to on-going and anticipated courses of action. Activity awareness is not merely a matter of coordinating state information. It must be continually negotiated and constructed throughout the course of a collaborative interaction. It is a process that is constitutive of collaboration.

### 16.8.2 Additional References

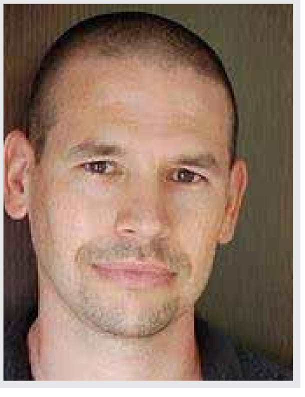
- Carroll, J.M., Neale, D.C., Isenhour, P.L., Rosson, M.B. & McCrickard, D.S. 2003. Notification and awareness: Synchronizing task-oriented collaborative activity. *International Journal of Human-Computer Systems 58*, 605-632.

- Carroll, J.M., Rosson, M.B., Convertino, G & Ganoe, C. 2006. Awareness and teamwork in computer-supported collaborations. *Interacting with Computers, 18,* 21-46.
- Carroll, J.M., Borge, M., Ganoe, C.H. & Rosson, M.B. 2011. Articulating collaborative contributions to activity awareness. In E. Salas, S. Fiore, & M. Letsky (Eds.), *Theories of Team Cognition: Cross-Disciplinary Perspectives,* New York: Routledge, pages 209-242.
- Carroll, J.M., Rosson, M.B., Farooq, U., & Xiao, L. 2009. Beyond being aware. *Information and Organizations, 19,* 3, 162-185.
- Dourish, P. and Bellotti, V. 1992. Awareness and Coordination in Shared Workspaces. *Proceedings of the ACM Conference on Computer-Supported Cooperative Work CSCW'92* (Toronto, Ontario), 107-114. New York: ACM.


## 16.9 COMMENTARY BY CLAY SPINUZZI

How to cite this commentary in your report

### Clay Spinuzzi



*© Clay Spinuzzi*

Associate professor of rhetoric at The University of Texas at Austin. Clay's research interests include research methods and methodology, workplace research, and computer-mediated activity. My first book, Tracing Genres through Organizations, was published by MIT Press in 2003 and was named NCTE's 2004 Best Book in Technical or Scientific Communication. My second book, Network, was publi...

Clay Spinuzzi
Clay Spinuzzi is a member of The Interaction Design Foundation

In Kaptelinin's conclusion, he argues that activity theory must develop to address new work organization. He says:

> "the conceptual framework of activity theory needs to be expanded to more adequately deal with coordination of multiple activities and cross-activity integration. Cross-activity integration is becoming an increasingly important issue in current uses of technology, characterized by complex social contexts (e.g., a combination of work and non-work factors typical of everyday practices of teleworkers) and employing multiple digital and non-digital technologies."

Examples might include university-industry partnerships (Gygi & Zachry 2010); massive multiplayer online role-playing games (Nardi 2010); coworking (Spinuzzi 2012, in press); classroom collaborations that span locations and disciplines (Paretti, McNair & Holloway-Attaway 2007); and sales engineers, who must bridge between clients and engineers (Ludvigsen et al. 2003). As Kaptelinin stated, such cross-activity work poses challenges to the conceptual framework of activity theory - and such examples are multiplying as activities become more networked.

Why is cross-activity integration such a critical issue now, and how must activity theory develop to address it? The answer lies, in part, in changes to work organization that were not anticipated during earlier stages of the theory's development. And the challenge lies in addressing these changes while keeping the theory relatively coherent.

The foundational ideas of activity theory came of age during the industrial era, grounded in Marx's critique of early industrialization (1990) and developed during the rapid industrialization of the Soviet Union (see especially Luria 1976). In fact, its early examples reflect agricultural and craft labor: hunting, fishing, farming, blacksmithing. But as Yrjö Engeström began developing third-generation activity theory (3GAT)[4], he recognized that work organization is changing in "the age of information technology" (1990, p.50), i.e., in the age of knowledge work, and that we are undergoing a historical transformation in the nature of expertise, moving toward "multi-professional team and network work and expertise" (1992, p.25). More recently, Engeström has suggested that we need a fourth generation of activity theory to address such work (2009, p.310). He argues that "Third-generation activity theory still treats activity systems as reasonably well-bounded, although interlocking and networked, structured units. What goes on between activity systems is processes, such as the flow of rules from management to workers". But, he says,

"In social production and peer production, the boundaries and structures of activity systems seem to fade away. Processes become simultaneous, multidirectional, and often reciprocal. The density and crisscrossing of

4. The term "third-generation activity theory" is controversial, since it suggests that activity theory has developed linearly from the first generation (Vygotsky) to a second (Leont'ev, Luria, Ilyenkov, and other Soviet thinkers), and finally to a third (Engeström and others primarily in the West). This is the story Engeström tells (1987 et passim) and it is told quite frequently by others operating in 3GAT (e.g., Lompscher 2006; Roth 2007). But others object that Engeström's version breaks theoretically and methodologically from its progenitors and that Engeström has tended to eclectically appropriate parts from theoretically divergent areas (Avis 2009; Bakhurst 2009; Martin & Peim 2009; Peim 2009; Witte 2005). Rather than engaging with that question, here, I restrict my scope to 3GAT, which I use as a synonym for Engeströmian AT.

processes makes the distinction between processes and structure somewhat obsolete. The movements of information create textures that are constantly changing but not arbitrary or momentary. ”

-- *Engeström 2009, p.309*

. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .

Like Kaptelinin and Engeström, others see challenges to activity theory as currently constituted (Bødker 2009; Lompscher 2006; Ruckriem 2009). For instance, Yamazumi (2009, p.212) argues that the knowledge society has shifted from mass production to interorganizational collaboration (cf. Castells 1996, 2003), resulting in “new types of agency [that] are collaborations and engagements with a shared object in and for relationships of interaction between multiple activity systems” (p.213). As Engeström puts it, “social production requires and generates bounded hubs of concentrated coordination efforts” (Engeström 2009, p.310), hubs in which interorganizational collaboration constitutes an aspect of the activity’s object (cf. Adler & Heckscher 2007; Gygi & Zachry 2010). Consequently, if we are to perform an activity theory analysis that is oriented toward knowledge work, we must examine the interorganizational collaborations to which they contribute.

Given these changes, activity theorists are increasingly concerned with addressing knowledge work. In the past few years, at least three collections on activity theory have addressed how it must adapt to discussing knowledge work (Sawchuk et al. 2006; Sannino et al 2009; Daniels et al. 2010), as have various monographs (Kaptelinin & Nardi 2006; Engeström 2008; Spinuzzi 2008).

As Kaptelin and Nardi argue: “Work itself is changing. Work is more distributed, more contingent, less stable. How do we understand social forms such as networks and virtual teams that partially replace standard organizational hierarchies? ... Knowledge work usually involves multitasking and working with diverse groups and individuals" (2006, p.26). And they describe the theoretical difficulties associated with this sort of work:

> "Can activity theory provide an account of multiple activities including an understanding of their structure and dependencies? In principle, yes. However, the conceptual apparatus of activity theory currently does not provide an elaborated set of concepts for the analysis of multiple activities."
>
> -- *Kaptelinin & Nardi 2006, p.256*

So the issue is known, but the elaborated concepts are yet to be developed. As we attempt to develop them, our great challenge will be to keep the theory coherent and focused while expanding it to address such analyses.

### 16.9.1 Additional references


- Adler, P. S., & Heckscher, C. (2007). Towards Collaborative Community. In C. Heckscher & P. S. Adler (Eds.), *The Firm as a Collaborative Community: Reconstructing Trust in the Knowledge Economy* (pp. 11-105). New York: Oxford University Press.
- Avis, J. (2009). Transformation or transformism: Engestrom's version of activity theory? *Educational Review*, *61*(2), 151-165.
- Bakhurst, D. (2009). Reflections on activity theory. *Educational Review*, *61*(2), 197-210.
- Bodker, S. (2009). Past experiences and recent challenges in participatory design research. In A. Sannino, H. Daniels, & K. D. Guti-

errez (Eds.), *Learning and expanding with activity theory* (pp. 274-285). New York: Cambridge University Press.

- Castells, M. (1996). *Rise of The Network Society* (p. 481). Malden, MA: Blackwell.
- Castells, M. (2003). *The Internet Galaxy: Reflections on the Internet, Business, and Society*. New York, NY: Oxford University Press.
- Daniels, H., Edwards, A., Engeström, Y., & Ludvigsen, S. R. (Eds.). (2010). *Activity Theory in Practice: Promoting Learning Across Boundaries and Agencies*. New York: Routledge.
- Engeström, Y. (1987). *Learning by expanding: An activity-theoretical approach to developmental research*. Helsinki: Orienta-Konsultit Oy. Retrieved from http://lchc.ucsd.edu/mca/Paper/Engestrom/expanding/toc.htm
- Engeström, Y. (1990). *Learning, working, and imagining: Twelve studies in activity theory*. Helsinki: Orienta-Konsultit Oy.
- Engeström, Y. (1992). *Interactive expertise: Studies in distributed working intelligence*. Helsinki: University of Helsinki.
- Engeström, Y. (2009). The future of activity theory: A rough draft. In A. Sannino, H. Daniels, & K. Gutierrez (Eds.), *Learning and expanding with activity theory*. New York: Cambridge.
- Gygi, K., & Zachry, M. (2010). Productive tensions and the regulatory work of genres in the development of an engineering communication workshop in a transnational corporation. *Journal of Business and Technical Communication*, *24*(3), 358-381.
- Kaptelinin, V., & Nardi, B. A. (2006). *Acting with technology: Activity theory and interaction design*. Cambridge, MA: MIT Press.
- Lompscher, J. (2006). The Cultural-Historical Development of Activity Theory: Some Aspects of Development. In P. Sawchuk, N.

Duarte, & M. Elhammoumi (Eds.), *Critical Perspectives on Activity: Explorations Across Education, Work, and Everyday Life* (pp. 35-51). New York: Cambridge University Press.

- Ludvigsen, S. R., Havnes, A., & Lahn, L. C. (2003). Workplace learning across activity systems: A case study of sales engineers. In T. Tuomi-Gröhn & Y. Engeström (Eds.), *Between school and work: New perspectives on transfer and boundary-crossing* (pp. 291-310). Boston: Pergamon.
- Martin, D., & Peim, N. (2009). Critical perspectives on activity theory. *Educational Review*, *61*(2), 131-138.
- Marx, K. (1990). *Capital: Volume 1*. New York: Penguin Books.
- Nardi, B. A. (2010). *My Life as a Night Elf Priest: An Anthropological Account of World of Warcraft*. Ann Arbor, MI: University of Michigan Press.
- Paretti, M. C., McNair, L. D., & Holloway-Attaway, L. (2007). Teaching Technical Communication in an Era of Distributed Work: A Case Study of Collaboration Between U.S. and Swedish Students. *Technical Communication Quarterly*, *16*(3), 327-352.
- Peim, N. (2009). Activity theory and ontology. *Educational Review*, *61*(2), 167-180.
- Roth, W.-M. (2007). Emotion at Work: A Contribution to Third-Generation Cultural-Historical Activity Theory. *Mind, Culture, and Activity*, *14*(1), 40-63.
- Ruckriem, G. (2009). Digital Technology and Mediation: A Challenge to Activity Theory. In A. Sannino, H. Daniels, & K. D. Gutierrez (Eds.), *Learning and Expanding with Activity Theory* (pp. 88-111). New York: Cambridge University Press.

- Sannino, A., Daniels, H., & Gutierrez, K. D. (Eds.). (2009). *Learning and Expanding with Activity Theory*. New York: Cambridge University Press.
- Sawchuk, P., Duarte, N., & Elhammoumi, M. (Eds.). (2006). *Critical Perspectives on Activity: Explorations Across Education, Work, and Everyday Life*. New York: Cambridge University Press.
- Spinuzzi, C. (2008). *Network: Theorizing knowledge work in telecommunications*. New York: Cambridge University Press.
- Spinuzzi, C. (2011)[5]. Losing by Expanding: Corralling the Runaway Object. *Journal of Business and Technical Communication*, *25*(4).
- Spinuzzi, C. (2012, in press). Working Alone, Together: Coworking as Emergent Collaborative Activity. *Journal of Business And Technical Communication*, *26*(4).
- Witte, S. P. (2005). Research in Activity: An Analysis of Speed Bumps as Mediational Means. *Written Communication*, *22*(2), 127-165.
- Yamazumi, K. (2009). Expansive Agency in Multi-Activity Collaboration. In A. Sannino, H. Daniels, & K. D. Gutierrez (Eds.), *Learning and Expanding with Activity Theory* (pp. 212-227). New York: Cambridge University Press.


### 16.9.2 Notes

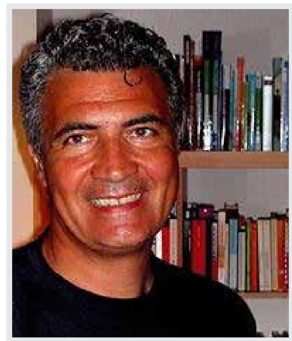
This commentary is a shorter version of the argument in Spinuzzi (2011)

## 16.10 COMMENTARY BY ANTONIO RIZZO

How to cite this commentary in your report

5. This argument is a shorter version of the argument in Spinuzzi (2011).

## Antonio Rizzo





Antonio Rizzo is Professor at the Communication Science Department, University of Siena. Previously, Antonio has been the Director of the Academy of Digital Arts and Science – ArsNova - in Siena and Chair of the European Association of Cognitive Ergonomics. He has also been Head of the Human Factors Division of the Italian Railways....

Antonio Rizzo
Antonio Rizzo is a member of The Interaction Design Foundation

### 16.10.1 On Mediation and Play

The socio-cultural Russian approach to human psychology has been the first tradition to deserve special attention and a specific role for tools in human cognition. Contrary to activity theory, the so far dominant approach to cognition, the representational theory of mind, assumes that tools have no inherent meaning or intrinsic role in human cognition. In fact, all computational models of the human mind share the core assumption that the use of a tool (e.g., a hammer) requires the extraction of sensory information about object properties (heavy, rigid, etc.), which can then be translated directly or indirectly into appropriate motor outputs (grasping, hammering, etc.). Already back in 1890, William James pointed out the paradox of this position, i.e. that to perceive properties of objects we need to know in advance what is the object for:

> "All ways of conceiving a concrete fact, if they are true ways at all, are equally true ways. *There is no property* ABSOLUTELY *essential to any one thing*. The same property, which figures at the essence of a thing on one occasion becomes a very inessential feature upon another. Now that I am writing, it is essential that I conceive my paper as a surface for inscription. If I failed to do that, I should have to stop my work. But if I wished to light a fire, and no other materials were by, the essential way of conceiving the paper would be as combustible material; and I need then have no thought of any of its other destinations. It is really all that it is: a combustible, a writing surface, a thin thing, a hydrocarbonaceous thing, a thing of eight inches one way and ten another, a thing just furlong east of a certain stone in my neighbor's field, an American thing, etc., etc., ad infinitum ... My thinking is first and last and always for the sake of my doing."
>
> *-- (James, 1890/2007, p. 333, italics and capitals in the original)*

We have to take into account that so far there is no clear experimental evidence supporting both the indirect (the semantic hypothesis) or direct (the cognitive definition of affordance) route between perception (vision/touch) and action (Osiurak at al., 2010). An alternative vision of the relationship between perception and action has received more empirical support (Adolph, Eppler, & Gibson, 1993).

Given this premise, it is not surprising that cognitive psychology has had a very limited impact on interaction design and that the original enthusiasm for a potential contribution to the design of “information processing” artifacts (Carroll, 1989) soon disappeared following the evidence that cognitive psychology had very little or no impact at all on design practices. As a result, the relationship between psychology and interaction design was best described as a relationship for mutual opportunity to learn (Carroll, 1991).

As opposed to representational theories of the human mind, soviet psychology attributes a special role to tools. Tools are, , as remarked by Victor Kaptelinin, integral to a fundamental feature of the socio-cultural approach, namely the process of mediation. And it is specifically on mediation that I would like to elaborate more Victor Kaptelinin’s description and report on my own experience of the role of mediation in the design process.

### 16.10.2 Mediation and its genesis

Mediation is a central aspect in Vygotsky’ s human psychology - constantly present in all his dynamic theoretical elaborations. The specific role that Vygotsky assign to tools can be summarized in this principle: human mental processes can be understood only if we understand the tools and signs that mediate these mental processes. Yet, in order to understand the mediation process we have to consider its relationship with the other two main topics of Vygotsky’s approach, that is the social genesis of human cognition and the developmental (genetic) method (Wertsh, 1985).

The starting point of my argument is the difference between mediated and non-mediated activities - a distinction that is related to the difference between elementary and higher mental functions: The central characteristic of elementary functions is that they are totally and directly determined by stimulation from the environment. For higher functions, the central feature is self-generated stimulation, that is the creation and use of artificial stimuli, which become the immediate causes of behavior (Vygotsky, 1978; p.39)

Mediation is usually presented as a way of transmitting existing cultural knowledge, here I will argue in favour of the other role of mediation (cf. Bødker & Klokmose, 2011), namely that of producing new meanings (and thus knowledge) by means of transforming the objects we interact with; the creation and use of artificial stimuli. This role I trace back to the seminal work of Vygotsky on children's play (1933/1982).

It is worthwhile to note that Vygotsky - in order to present his ideas about children's play - takes his point of departure in a very early idea of affordance, that of Kurt Lewin's valence. Vygotsky quotes a study carried out by Lewin where it is shown how very young children in the attempt to exploit the opportunities for action offered by a stone exhibit a behavior that is strongly determined by the conditions in which the activity takes place.

In the following videos we can observe the original recording of the Kurt Lewin's study. The first video presents the interaction with the stone by Hannah (19 months old), while the second video shows the performance of Han (who is older than Hannah).

Video 1: Hannah is one year and seven months old. The stone has a positive valence in the momentary living space of the child. The child is attracted by the stone. In order to sit down, the child has to turn around, that is away from the goal. This detour to reach the goal is extremely difficult for children.

Video 2: Hans solves the problem in an intelligent fashion. He does not lose sight of his goal

For Vygotsky the interaction exhibited by the two children and the description provided by Kurt Lewin is a real illustration of the extent to which a very young child is bound in her/his action by situational constraints. He states:

> "It is hard to imagine a greater contrast to Lewin's experiment showing the situational constraints on activity than what we observe in play. It is here that the child learns to act in a cognitive, rather than an externally visual, realm by relying on internal tendencies and motives and not on incentives supplied by external things. ...Lewin concludes that things dictate to the child what he must do: a door demands to be opened and closed, a staircase to be run up, a bell to be rung. In short, things have an inherent motivating force in respect to a very young child's actions and determine the child's behavior to such an extent that Lewin arrived at the notion of creating a psychological topology "

*-- (Vygotsky, 1978; p. 96)*

Instead, for Vygotsky:

> “in play, things lose their motivating force. *The child sees one thing but acts differently in relation to what he sees*. …in play the child creates the structure meaning/object, in which the semantic aspect – the meaning of the thing – dominates and determines his behavior. To a certain extent meaning is freed from the object with which it was directly fused before. I would say that in play a child concentrates on meaning severed from objects, but that it is not severed in real action with real objects. ”

*-- (Vygotsky, 1978; p. 96 – 97; italics in the original)*

However, here comes an interesting consideration, namely that in the genesis the separation is not totally arbitrary:

> “This is not to say that properties of things as such have no meaning. Any stick can be a horse, but, for example, a postcard can never be a horse for a child. Goethe’s contention that in play any thing can be anything for a child is incorrect. Of course, for adults who can make conscious use of symbols, a postcard can be a horse. If I want to show the location of something, I can put down a match and say, “This is a horse.” And that would be enough. For a child it cannot be a horse: one must use a stick. Therefore, this is play, not symbolism. A symbol is a sign, but the stick is not the sign of a horse. Properties of things are retained, but their mean-

ing is inverted, i.e., the idea becomes the central point. It can be said that in this structure things are moved from a dominating to a subordinate position. ”

*-- (Vygotsky, 1978; p. 98)*

Indeed, there is experimental evidence in social play to support that meaning is understood not by the shape, colour or other features of the objects involved in the activity but by the actions the object allows to be performed (Szokolsky, 2006)

It's the action pattern that provides the cue for what is intended, not the objects.Indeed there is evidence that very young children (12 – 18 months old) imitate significantly more often when the pattern of action performed by an adult involves (is mediated by) an object compared to a condition where the same pattern of action is executed without any object (Rizzo and Carnesecchi, 2011). Pretend play is a privileged way of staying in touch with the environment as well as stepping out of the environment to mentally modify it.

For me, the more dramatic example of this was provided by the children involved in the design of POGO world (Rizzo et al 2003; Decortis and Rizzo, 2002).


Figure 16.1: Pogo Mosaic. A collage of snapshot of pogo prototyping from version 1 (up left) to version 2 (bottom right).

Together with Françoise Decortis and Patrizia Marti we were in charge of the mock-upping and testing (role prototyping through dramatization) of the design concepts produced by the Domus Academy. And what we observed was that kids had no problems at all in overruling the intended use of the mock-up and produce new opportunities for actions and relationship with existing objects.

In the POGO dramatization we observed how childrens' behaviour was guided by a merge between the sensory-motor affordance of the mock-ups and the meaning of the current situation. The same object, for example the pogo torch (a device to capture and project sounds and images), was used as a way to talk to yourself in one situation or as a way to move very large objects in another situation according to the meaning the children were negotiated in their play.

Video 3: A short clip of different situtations where children use the first generation of POGO mock-ups

This resulted in new functionalities that were not anticipated by the design team. The attempt to give new objects with specific functionalities to children in order to see what role the objects would play in their activity did not work. As the children did not get directions by a teacher, they appropriated the tool to the game they played, thus producing new meanings on the fly. These new meanings were however linked to the actions (movements) made feasible by the objects.

It was pretty clear that we, the designers, were thinking about functions and they, the children, where negotiating and producing meaning through their activity: Meaning comes first, function later.

An interesting observation was that new opportunities for action introduced by the children (that is, specific manipulations of the torch or of the mambo) had a deontic power that sensorimotor affordances did not have: Children imposed the new opportunities for action onto their peers: “Nooo, it is not that way... look!” A situation never observed for any sensorimotor affordance (ways of grasping, pushing or waving).

Such observations inspired me to look for theoretical elaborations of the concept of affordance, which led me to Michael Tomasello’s idea of intentional affordances and their role in children’s cultural learning. Tomasello notes that children are involved in intentional mirroring process (imitation and in some sense emulation) and through these processes, the children start to perceive objects and artifacts as elements that evoke a set of affordances, beyond basic sensory-motor affordances:

> “Such affordances rest upon the understanding of the intentional relations that other persons have with that object or artifact-that is, the intentional relations that other person have to the world through the artifact.”
>
> -- *(Tomasello, 1999, pp 84-85)*

This way of producing and sharing new objects’ attributes had a profound impact on the design of POGO and subsequently influenced the whole design strategy:

> “I) POGO world was built on the idea to allow composition/recomposition of existing tools, and to promote construction/deconstruction of new tools by the children and teachers

II) The development of prototypes embodied the same approach: the tools were built through a strategy named "Smart Shopping": deconstructing existing hardware and software tools (joystick, console, screen, cameras, memory-card, rfid, editing software, file management systems) and constructing POGO tools."

*-- Rizzo and Rutgers, 2004, p 3.*

Most of this was done in playful sessions involving designers, psychologists, and teachers (and sometime selected children) where a mix of mock-ups, existing objects, semi-working prototypes were put on stage following a hint script and improvisational theatre techniques (Rizzo and Bacigalupo, 2004). All the materials were used to explore new territories of interactions and the ease of which the materials propagated among the team members was also tested.

A few year later, Banzi, the team leader and inventor of Arduino at Ivrea, made a similar point (although not mentioning the social dimension) by introducing the term Tinkering to the interaction design community (a term originally coined by Francois Jacob in 1977 in biology):

"We believe that it is essential to play with technology, exploring different possibilities directly on hardware and software - sometimes without a very defined goal. Reusing existing technology is one of the best ways of tinkering. Getting cheap toys or old discarded equipment and hacking them to make them do something new is one of the best ways to get great results"

*-- Banzi, 2009*

One of the best definitions of tinkering, as also acknowledged by Banzi, is the definition provided on the former website of the Exploratorium Museum in San Francisco:

> "Tinkering is what happens when you try something you don't quite know how to do, guided by whim, imagination, and curiosity. When you tinker, there are no instructions-but there are also no failures, no right or wrong ways of doing things. It's about figuring out how things work and reworking them. Contraptions, machines, wildly mismatched objects working in harmony- this is the stuff of tinkering."

Tinkering is, at its most basic, a process that marries play and inquiry.

### 16.10.3 Conclusion

I believe that to "tap into" the heuristic power of Activity Theory we need more analytical tools that, on the one hand,impactdesign processes, and, on other hand, may have an impact directly on the artifacts we design. To my mind, there is plenty of room at the interplay between sensorimotor and intentional affordances (Rizzo et al, 2009); a room that combines the dynamic and evolving relationship between non-mediated and mediated action, a room that needs to be explored and exploited in the design of human interaction with her/his environment.

## 16.10.4 Additional References


- Adolph, K. E., Eppler, M. A., & Gibson, E. J. (1993). Development of perception of affordances. In C. Rovee-Collier & L. P. Lipsett (Eds.), *Advances in infancy research* Vol. 8, Norwood, NJ: Ablex. pp. 51-98
- Banzi, M. (2008) *Getting Started with Arduino*. O'Reilly/Make Books
- Carroll, J. (1987) *Interfacing thought: Cognitive aspects of human-computer interaction*. Cambridge, MA Bradford Books/MIT Press.
- Carroll, (1991) *Designing Interaction: Psychology at the Human-Computer Interface*. Cambridge, MA: Cambridge University Press
- Decortis, F. & Rizzo, A. (2002). New active tools for supporting narrative structures. *Personal & Ubiquitous Computing*. 6 : pp. 416–429
- James, W. (2007). *The principles of psychology, Vol. 2*. New York: Cosimo Classics. (Original work published 1890)
- Osiurak, F.; Jarry, C., Le Gall, D. (2010) Grasping the affordances, understanding the reasoning: Toward a dialectical theory of human tool use. *Psychological Review*, Vol 117(2), Apr 2010, pp. 517-540.
- Rizzo, A. and Carnesecchi, M. (2011) Il ruolo euristico degli oggetti nell'apprendimento sociale (The heuristic role of objects in social learning). Paper presented at the XXIV Congresso Nazionale dell A.I.P. Genova 19 -21 Settembre, Italy.
- Rizzo, A. and Rutgers, J. (2004) The Construction/Deconstruction of Ambient
  CHI 2004 Workshop Eds by Anton Nijholt, Thomas Rist, Kees Tuijnenbreijer Lost in Ambient Intelligence? 24-29 April, Vienna, Austria
- Rizzo, A., Del Monte, M., Rubegni, E., Torsi, S. (2009). The interplay between sensory-motor and intentional affordances. In Work-

shop on Children & Embodied Interaction. Ed by Alissa Antle at IDC09. 3 – 5 June, Como, Italy.

- Rizzo, A. & Bacigalupo, M. (2004) Scenarios: Heuristics for action. In: *Proceedings of XII European Conference on Cognitive Ergonomics, Living and Working with Technology*, ed. D. J. Reed, G. Baxter & M. Blythe, EACE pp. 153-160.
- Rizzo, A. Marti, P. Decortis, F., Moderini, C., Rutgers, J. (2003) The design of POGO story world. In Hollnagen E. *Handbook of Cognitive Task Design*. Mahwah, NJ: Lawrence Erlbaum, 2003. pp.577-602.
- Szolonsky, A. (2006). Object use in Pretend Play: Symbolic or Functional? In Alan Costall and Ole Dreier (Eds.) *Doing Things with Things*. Ashgate Publishing. Pp. 67–85.
- Tomasello, M. (1999) *The cultural origins of human cognition*. Cambridge MA Harvard University Press.
- Vygotsky, L. (1978) Mind in Society. The development of higher psychological processes. Cambridge MA. Harvard University Press (Original work published 1933)
- Wertsh, J.V. (1985) Vygotsky and the social formation of mind. Cambridge MA Harvard University Press.

## 16.11 COMMENTARY BY STEPHEN VOIDA

How to cite this commentary in your report

### Stephen Voida

*© Stephen Voida*

Stephen Voida is a lecturer and research scientist in the Department of Informatics at the University of California, Irvine. He has a Ph.D. in Computer Science, an M.S. in Human–Computer Interaction, and a certificate in Cognitive Science from the Georgia Institute of Technology and a B.S. in Computer Science from Arizona State University. His research explores the development of novel u...

Stephen Voida
Stephen Voida is a member of The Interaction Design Foundation

In this chapter (as well as in numerous previous books and articles), Kaptelinin provides a thoughtful and comprehensive review of Activity Theory, its history, and some of the many ways that the conceptual framework has been taken up and appropriated by the CHI and interaction design community.

Whether explicitly acknowledged as such or not, Activity Theory has had a significant impact on the way that researchers and practitioners have approached the design and evaluation of interactive systems in the mobile and ubiquitous computing era. Bannon and Bødker introduced Activity Theory to the interaction design community at around the same time that Weiser published his seminal article establishing the vision for ubiquitous computing (Weiser 1991). While this wasn't a coordinated or intentional effort, both researchers were responding in their own way to the limitations of computing technology and the way that we conceptualized people's relationships with computers at the time. The desktop computing paradigm of the early 1990s placed practical limitations on the contexts in which human–computer interaction could occur, but the movement towards making computers smaller, more mobile, and more often embedded into other objects made it clear that computational tools would soon permeate the everyday world and play an much more significant role in all kinds of human activities. Activity Theory broke from the established theories of interaction-as-dialog and cognition-as-information processing to provide a lens for understanding how humans might interact with ubiquitous computational technologies in a much greater breadth of contexts beyond number crunching and word processing in the workplace.

One of the most striking things about the relationship between Activity Theory and HCI is the framework's continued success and longevity as a relevant way of thinking about the mediating role of computational tools in the face of a dynamic and rapidly evolving technological landscape. Not only has Activity Theory been adopted as a general-purpose analytic tool within HCI (cf. section 16.3.4, above), the conceptual framework has been extended to better support reflection about the temporality and interconnectness of activities in knowledge work organizations (Boer, van Baalen & Kumar 2002) and to incorporate the notion of external environmental factors (Döweling, Schmidt & Göb 2012). It has also been used as the basis for new methodologies that aim to make sense of empirical data collection carried out in complex, collaborative work environments, such as hospitals (Bardram & Doryab 2011).

But perhaps the most significant re-purposing of Activity Theory has been in re-casting what was primarily an analytic, inspirational, and discursive tool to one that has served as a guidepost in both the design and implementation of interactive systems. While the early command-line and windowed GUI interface paradigms were largely focused on supporting the creation or manipulation of a single file, document, or electronic artifact at a time, Activity Theory challenged the premise that computational support should focus on interaction with a single, decontextualized document at a time. Activity Theory's emphasis on articulating the dynamic, at times complex, and occasionally conflicting relationships among subjects, tools/artifacts, and social/environmental context has both influenced the structure of various personal information management and desktop interfaces (including those enumerated by Kaptelinin in section 16.3.5), as well as the underlying data representations that are used to organize electronic artifacts and support further exploration of what has become known as the "activity-based" or "activity-oriented" computing movement.

In my research, I have found that the theoretical framework provided by Activity Theory-and particularly the modern instantiations articulated by Engeström (1987) and Boer et al. (2002)-align and resonate surprisingly well with empirical observations of the ways that information workers organize their workspaces (e.g., Malone 1983), the ways that they transform themselves in the process of carrying out information work (e.g., Kidd 1994), and how they handle transitions among and interruptions within ongoing activities throughout the work day (e.g., González & Mark 2004). In the systems that I built to support information work and explore the role of activity in interface design, these points of resonance helped to shape both the systems' interface design and their underlying data structures. The Kimura system (MacIntyre et al. 2001) displayed interactive visualizations of ongoing activities on an electronic whiteboard to facilitate multitasking and activity awareness. Each of the visualizations brought together representations of the computer application windows (*mediating tools*) that had been used over the course of the activity, along with icons representing the people (*community*) with whom further collaboration would be required in order to bring the activity to a successful conclusion. The Giornata system (Voida, Mynatt & Edwards 2008) extended Kimura's model of activities to include discrete electronic resources and broadened the system's focus from primarily supporting multitasking to also facilitate collaboration and evolving personal information management practices. The data structure behind each activity in Giornata encoded a user-generated series of tags describing the current goals or meaning of the activity (which could change over time); a flexible set of documents and applications, both live and archival, representing the computational tools used to mediate, transform, and generate information content; and a "palette" of contact icons allowing quick access to and information sharing with the other people associated with the activity. Over the course of this research, I identified a number of key challenges that are brought to the fore when an activity-theoretical perspective is used in the design process for both desktop and ubiquitous computing systems (Voida, Mynatt & MacIntyre 2007, Voida 2008). Like Kaptelinin, Nardi & Macaulay's activity checklist (1999), these challenges serve as scaffolding to transition between various facets of the Activity Theory framework and the articulation of concrete system requirements.

Even though Activity Theory has been part of the theoretical tool belt in HCI for nearly two decades, there are still areas in which the flexibility of the framework raises practical issues about how to apply its concepts most effectively. For example, the hierarchical structure of activities and the inherent variability in the granularity at which people describe-and organize-their ongoing activities sometimes makes it difficult to adequately model the relationships at play in empirical data, especially across multiple individuals or multiple activities. Although an Activity

Theory analysis may be carried out at any of the levels of the hierarchy (e.g., operation, action, activity, or aggregate/higher-level activity), different people tend to articulate their activities at different levels of detail, depending on, for example the scope (i.e., complexity, anticipated duration, or importance) of the activity, the person's role in the activity, and the perceived expertise or familiarity level of the interviewer/listener. Likewise, computational systems that aim to explicitly model activities as part of the user experience often cannot anticipate at what level of detail individuals might wish to represent their work-for example, "editing a revision of a chapter" versus "writing a book" (Voida, Mynatt & MacIntyre 2007). These differences sometimes make it difficult to anticipate what specific types of computational support might be appropriate or helpful at a given time. Furthermore, facilitating collaboration through computational activity representations when participants have created representations of their work at different levels of granularity can be problematic.

One of the relatively underutilized aspects of Leontiev's framework in interaction design is his focus on the continual development of activity systems. Historically, most personal computing systems have been designed to represent the state of a data structure at a particular point in time (the present). Representations of temporality have tended to exist as simple linear state-management tools (e.g., *undo* and *redo*) or very formal representation of milestones and revision numbers, such as those found in version control and transaction-based database systems. In most conventional operating systems and mobile computing platforms, users must maintain their own representations of information-through-time, constructing and maintaining their own artifact histories, often by duplicating documents at each milestone or using auxiliary information systems (like e-mail) to archive the state of a document through multiple points in time. Providing better solutions for these actions is becoming ever more urgent as computational systems enable the creation and sharing of more and more content through an increasing diversity of platforms and online services. If our computational systems were to better reflect the activity-theoretical idea that activities continually develop and evolve, we might arrive at better support for information management, long-term information curation, externalization, and routinization, but the interaction techniques for dealing with this kind of temporality are neither widely utilized in mainstream computing systems nor will they be as familiar as the typical application- and document-centric interaction paradigm.

### 16.11.1 Additional References


- Bardram, Jakob and Doryab, Afsaneh (2011):
  Activity analysis: applying activity theory to analyze complex work in hospitals. In: *Proceedings of ACM CSCW11 Conference on Computer-Supported Cooperative Work*. March 19–23, 2011, Hangzhou, China (pp. 455–464). Available online
- Boer, Niels-Ingvar, van Baalen, Peter J., and Kumar, Kuldeep (2002):
  An activity theory approach for studying the situatedness of knowledge sharing. In: *Proceedings of the 35th Hawaii International Conferences on System Sciences*. January 7–10, 2002, Big Island, Hawaii.
- Döweling, Sebastian, Schmidt, Benedikt, and Göb, Andreas (2012):
  A model for the design of interactive systems based on Activity Theory. In: *Proceedings of the ACM CSCW12 Conference on Computer Supported Cooperative Work*. February 11–15, 2012, Bellevue, Washington (pp. 539–548).
- González, Victor M. and Mark, Gloria (2004): "Constant, constant, multi-tasking craziness":
  Managing multiple working spheres. In: Dykstra-Erickson, Elizabeth and Tscheligi, Manfred (eds.) *Proceedings of ACM CHI 2004 Conference on Human Factors in Computing Systems*. April 24–29, 2004, Vienna, Austria (pp. 113–120). Available online

- Kidd, Alison (1994):
  The marks are on the knowledge worker. In: Adelson, Beth, Dumais, Susan and Olson, Judith S. (eds.), *Proceedings of the ACM CHI 94 Human Factors in Computing Systems Conference*. April 24–28, 1994, Boston, Massachusetts (pp. 186–191).
- Malone, Thomas W. (1983):
  *How Do People Organize Their Desks? Implications for the Design of Office Information Systems*. In:
  ACM Transactions on Information Systems, 1 (1) pp. 99-112.
- MacIntyre, Blair, Mynatt, Elizabeth D., Voida, Stephen, Hansen, Klaus Marius, Tullio, Joe and Corso, Gregory M. (2001):
  Support for multitasking and background awareness using interactive peripheral displays. In:
  Marks, Joe and Mynatt, Elizabeth D. (eds.), *Proceedings of the 14th Annual ACM Symposium on User Interface Software and Technology*. November 11–14, 2001, Orlando, Florida (pp. 41–50).
  Available online
- Voida, Stephen (2008): Exploring user interface challenges in supporting activity-based knowledge work practices.
  Doctoral dissertation. Atlanta, Georgia, Georgia Institute of Technology.
- Voida, S., Mynatt, Elizabeth D., and MacIntyre, Blair (2007): Supporting activity in desktop and ubiquitous computing. In:
  Kaptelinin, Victor and Czerwinski, Mary (eds.), *Beyond the desktop metaphor:*
  *Designing integrated digital work environments*. Cambridge, Massachusetts: MIT Press (pp. 195–222).
- Weiser, Mark (1991): The computer for the 21st century. In:
  *Scientific American 265* (3), pp. 94–104.

## 16.12 COMMENTARY BY KLAUS B. BAERENTSEN

How to cite this commentary in your report

**Klaus B. Baerentsen**

Klaus Bærentsen is associate professor at the Department of Psychology, University of Aarhus. His research area concerns human consciousness in a broad sense, encompassing evolution, cultural history and ongoing individual life activity, as well as the brain activity supporting conscious experience. His main theoretical inspirations stems from a combination of the Russian theory of Activ...

Klaus B. Baerentsen
Klaus B. Baerentsen is a member of The Interaction Design Foundation

Klaus is working hard on his commentary - please stay tuned!

## 16.13 COMMENTARY BY ELLEN CHRISTIANSEN

How to cite this commentary in your report

### Ellen Christiansen

© *Ellen Christiansen*

Ellen Christiansen is Professor at Aalborg University. She works with activity theory within the humanistic computer science studies, concentrating on the communicative aspects of systems development, especially in relation to artificial intelligence...

Ellen Christiansen
Ellen Christiansen is a member of The Interaction Design Foundation

Designers need tools for understanding the context of whatever new artifact they are working on. Certainly, among academically trained interface designers, activity theory has been instrumental to such understanding since Bødker's book *'Through the Interface'* came out. It was, as pointed out by Kaptelinin in this chapter, the very same year (1991) Carroll, in a book with the title *'Designing Interaction'* moved the HCI agenda from 'interface' to 'interaction', while pointing out that for this purpose: designing *interaction*, cognitive psychology alone fell short as a foundation understanding context.

Bødker's book *'Through the Interface'* established the 'subject-instrument-object'-relationship as an indispensible syntax for legitimate sentences in the field of HCI. She made it clear that the interface is not an endpoint, but a window to a world of activity, which people are inclined to embark on anyways. Hence, the success of an interface depends on the degree to which the interaction brings this world closer. Since then, humans have come to experience interaction in an explosion of ways, on digital arenas for artwork, gaming, and amusements of all sorts, to the extent that today we have a hard time telling what, regarding interaction, is text, and what is context, what is work and what is leisure.

Hence, when, in the field of HCI, activity theory have matured to take position between what Kaptelinin calls "the visible landmarks of the theoretical landscape of Human-Computer Interaction", we may ask if activity theory is more than a memorial. From a practitioners' point of view, at least, we may ask, not what activity theory delivered to yesterday's HCI-designers and design thinking, but what it has to offer the designers of tomorrow.

Given that digital technology pervades all aspects of human life, almost as the air we breathe, it can be tempting to dismiss questions about context all together: Designers were never able to predict use, and who would know today, which needs to fulfill tomorrow? Soon it will be hard to distinguish a non-robotic human from a robotic one, Human-Computer Interaction may dissolve, interaction designers may work in global app-stores, and the only thing we know for sure is that we breathe, and that the stock market is a roller-coaster.

More persistent than these fluctuations, however, is the fact of the pendulum: what comes up, must go down. While, at this moment rational thought and critical reflection about context seems out of fashion, in the next moment, which may well be soon, designers will experience a craving for being able to reduce complexity of context in ways they can comprehend, communicate, criticize, and improve. To satisfy that hunger the concept of activity, and the models for analysis presented in Kaptelinin's chapter 16 on 'Activity Theory' will be a place to start to feed. Here you find key concepts for understanding the way humans interact with the world: 'tools', 'mediation', and 'development' in relation to 'action', 'activity' and 'operation', modeled in a hierarchical structure, a key rack, where to hang your experiences. Taken as a tool for designerly thinking, activity theory will help designers to communicate, sort out, categorize and evaluate experiences of any kind imaginable - definitely not an end-point, but possibly an access-point for communication about artifacts, of a trustworthy kind.Not only is Victor Kaptelinin extraordinarily well read in the research literature on how and why to apply activity theory, he is also sufficiently experienced as a design practitioner to know designers' needs regarding tools for thinking. Therefore, his account of activity theory in this chapter provides both a good blend of key-rack models as well as a scholarly grounding of the theory behind the key-racks.

## 16.14 REFERENCES

Bannon, *Liam (1991): From human factors to human actors: the role of psychology and human-computer interaction studies in system design. In:* Greenbaum, *Joan and* Kyng, *Morten (eds.). "Design at Work: Cooperative Design of Computer Systems". Hillsdale, New Jersey: Lawrence Erlbaum Associatespp. 25-44*

Bardram, *Jakob E.,* Bunde-Pedersen, *Jonathan and* Soegaard, *Mads (2006): Support for activity-based computing in a personal computing operating system. In:* Proceedings of ACM CHI 2006 Conference on Human Factors in Computing Systems *2006. pp. 211-220*

Basov, *Mikhail Ia. (1991): The Organization of Processes of Behavior (A Structural Analysis) (originally published in Russian in 1931). In* Journal of Russian and East European Psychology, *29 (5) pp. 14-83*

Bedny, *Gregory Z. and* Harris, *Steven Robert (2005): The Systemic-Structural Theory of Activity: Applications to the Study of Human Work. In* Mind, Culture and Activity, *12 (2) pp. 128-147*

Bedny, *Gregory and* Karwowski, *Waldemar (2003): A Systemic-Structural Activity Approach to the Design of Human-Computer Interaction Tasks. In* International Journal of Human-Computer Interaction, *16 (2) pp. 235-260*

Bertelsen, *Olav W. and* Bødker, *Susanne (2003): Activity Theory. In:* Carroll, *John M. (ed.). "HCI Models, Theories, and Frameworks". San Francisco: Morgan Kaufman Publisherspp. 291-324*

*Blackler, F. (1995): Activity theory, CSCW and organizations. In: Monk, Andrew F. and Gilbert, Nigel G. (eds.). "Perspectives on HCI: Diverse Approaches". Academic Press*

*Brushlinsky, Andrej V. and Aboulhanova-Slavskaya, Ksenia A. (2000): The historical context and modern denotations of the fundamental work by S. L. Rubinshtein. Afterword to the 4th edition of Foundations of General Psychology (in Russian). In: Rubinstein, Sergey L. (ed.). "Foundations of General Psychology - 4th edition". St. Petersburg, Russia: Piter*

*Bødker, Susanne (2006): When second wave HCI meets third wave challenges. In: Proceedings of the Fourth Nordic Conference on Human-Computer Interaction 2006. pp. 1-8*

*Bødker, Susanne (1989): A Human Activity Approach to User Interfaces. In Human-Computer Interaction, 4 (3) pp. 171-195*

*Bødker, Susanne (1991): Through the Interface - A Human Activity Approach to User Interface Design. Hillsdale, NJ, Lawrence Erlbaum Associates*

*Bødker, Susanne and Andersen, Peter B. (2005): Complex Mediation. In Human-Computer Interaction, 20 (4) pp. 353-402*

*Carroll, John M. (ed.) (1991): Designing Interaction: Psychology at the Human-Computer Interface. Cambridge, UK, Cambridge University Press*

*Carroll, John M. (2012). Human Computer Interaction - brief intro. Retrieved 6 June 2013 from [URL to be defined - in press]*

*Cole, Michael and Engeström, Yrjö (1993): A cultural-historical approach to distributed cognition. In: Salomon, Gavriel (ed.). "Distributed Cognitions: Psychological and Educational Considerations (Learning in Doing: Social, Cognitive and Computational Perspectives)". Cambridge University Press*

*Cooper, Geoff and Bowers, John (1995): Representing the user: notes on the disciplinary rhetoric of human-computer interaction. In: Thomas, Peter J. (ed.). "The Social and Interactional Dimensions of Human-Computer Interfaces (Cambridge Series on Human-Computer Interaction)". Cambridge University Press*

*CRADLE (2011). CRADLE, Center for Research on Activity, Development and Learning. Retrieved 19 April 2011 from CRADLE: http://www.helsinki.fi/cradle/*

*Cypher, Allen (1986): The structure of users' activities. In: Norman, Donald A. and Draper, Stephen W. (eds.). "User Centered System Design: New Perspectives on Human-Computer Interaction". Hillsdale, NJ: Lawrence Erlbaum Associatespp. 243-264*

Daniels, Harry, Edwards, Anne, Engeström, Yrjö, Gallagher, Tony and Ludvigsen, Sten R. (eds.) (2009): Activity Theory in Practice: Promoting Learning Across Boundaries and Agencies. Routledge

Dourish, Paul (2001): Where the Action Is: The Foundations of Embodied Interaction. MIT Press

Engeström, Yrjö (1987): Learning by expanding: An activity-theoretical approach to developmental research. Orienta-Konsultit Oy

Engeström, Yrjö (ed.) (1990): Learning, Working and Imagining. Orienta-Konsultit Oy

Engeström, Yrjö, Miettinen, Reijo and Punamaki, Raij (eds.) (1999): Perspectives on Activity Theory (Learning in Doing Social, Cognitive and Computational Perspectives). Cambridge University Press

Gay, Geraldine and Hembrooke, Helene (2004): Activity-Centered Design: An Ecological Approach to Designing Smart Tools and Usable Systems (Acting with Technology). The MIT Press

Halverson, Christine (2002): Activity Theory and Distributed Cognition: Or What Does CSCW Need to DO with Theories?. In Computer Supported Cooperative Work, 11 (1) pp. 243-267

Hassenzahl, Marc (2011). User Experience and Experience Design. Retrieved 6 June 2013 from [URL to be defined - in press]

Heidegger, Martin (1962): Being and Time. English translation 1962.

Hollan, James D., Hutchins, Edwin and Kirsh, David (2000): Distributed Cognition: Toward a New Foundation for Human-Computer Interaction Research. In ACM Transactions on Computer-Human Interaction, 7 (2) pp. 174-196

Ilyenkov, E. V. (2008): Dialectic Logic: Essays on Its History and Theory. Aakar Books

Jonassen, David H. and Rohrer-Murphy, Lucia (1999): Activity theory as a framework for designing constructivist learning environments. In Educational Technology Research & Development, 47 (1) pp. 61-79

Kaptelinin, Victor (2003): UMEA: translating interaction histories into project contexts. In: Cockton, Gilbert and Korhonen, Panu (eds.) Proceedings of the ACM CHI 2003 Human Factors in Computing Systems Conference April 5-10, 2003, Ft. Lauderdale, Florida, USA. pp. 353-360

Kaptelinin, Victor and Czerwinski, Mary (2007): Introduction. In: Kaptelinin, Victor and Czerwinski, Mary (eds.). "Beyond the Desktop Metaphor: Designing Integrated Digital Work Environments". The MIT Press

Kaptelinin, Victor and Nardi, Bonnie A. (2006): Acting with Technology: Activity Theory and Interaction Design. The MIT Press

Kaptelinin, Victor and Nardi, Bonnie A. (2012): Activity Theory in HCI: Fundamentals and Reflections. Morgan and Claypool

Kaptelinin, Victor, Nardi, Bonnie A., Bødker, Susanne, Carroll, John M., Hollan, James D., Hutchins, Edwin and Winograd, Terry (2003): Post-cognitivist HCI: second-wave theories. In: Cockton, Gilbert and Korhonen, Panu (eds.) Extended abstracts of the 2003 Conference on Human Factors in Computing Systems - CHI 2003 April 5-10, 2003, Ft. Lauderdale, Florida, USA. pp. 692-693

Kaptelinin, Victor, Nardi, Bonnie A. and Macaulay, Catriona (1999): Methods & tools: The activity checklist: a tool for representing the. In Interactions, 6 (4) pp. 27-29

Koschmann, Timothy, Kuutti, Kari and Hickma, Larry (1998): The Concept of Breakdown in Heidegger, Leont'ev, and Dewey and Its Implications for Education. In Mind, Culture, and Activity, 5 (1) pp. 25-41

Leontiev, Aleksei Nikolaevich (1978a): Activity, consciousness, and personality. Prentice-Hall

Leontiev, Aleksei Nikolaevich (1978b): Activity, consciousness, and personality (originally published in Russian in 1975). Prentice-Hall

Leontiev, Aleksei Nikolaevich (1981): Problems of the development of the mind (originally published in Russian in 1959). Moscow, Russia, Progress

Mescherjakov, Boris and Zinchenko, Vladimir P. (eds.) (2003): Big Psychological Dictionary (in Russian). Prime-Evroznak

Moran, Thomas P. (2006): Activity: Analysis, Design, and Management. In: Bagnara, Sebastiano and Smith, Gillian Crampton (eds.). "Theories and Practice in Interaction Design (Human Factors and Ergonomics Series)". Lawrence Erlbaum Associates

Moran, Thomas P., Cozzi, Alex and Farrell, Stephen P. (2005): Unified activity management: supporting people in e-business. In Communications of the ACM, 48 (12) pp. 67-70

Mwanza, Daisy (2002): Conceptualising work activity for CAL systems design. In J. Comp. Assisted Learning, 18 (1) pp. 84-92

Nardi, Bonnie A. (ed.) (1996a): Context and Consciousness: Activity Theory and Human-Computer Interaction. MIT Press

Nardi, Bonnie A. (1996b): Studying context: A comparison of activity theory, situated action models, and distributed cognition. In: Nardi, Bonnie A. (ed.). "Context and Consciousness: Activity Theory and Human-Computer Interaction". MIT Press

*Nardi, Bonnie A. (2005): Objects of desire: Power and passion in collaborative activity. In Mind, Culture, and Activity, 12 pp. 37-51*

*Norman, Donald A. (2005): Human-centered design considered harmful. In Interactions, 12 (4) pp. 14-19*

*Norman, Donald A. (1998): The Invisible Computer: Why Good Products Can Fail, the Personal Computer Is So Complex and Information Appliances Are the Solution. MIT Press*

*Quek, Amanda and Shah, Hanifa (2004): A Comparative Survey of Activity-Based Methods for Information Systems Development. In: ICEIS 2004 2004. pp. 221-232*

*Rabardel, Pierre and Bourmaud, Gaetan (2003): From computer to instrument system: a developmental perspective. In Interacting with Computers, 15 (5) pp. 665-691*

*Rogers, Yvonne (2004): New Theoretical Approaches for HCI. In Annual Review of Information Science and Technology, (38) pp. 1-43*

*Rubinshtein, Sergey L. (1986): The principle of creative self-activity (on the philosophical foundations of modern pedagogy). Originally published in 1922 in Russian. In Voprosy Psikhologii, 4 pp. 101-107*

*Rubinshtein, Sergey L. (1946): Foundations of General Psychology. Second edition. (in Russian). Moscow, Russia, Uchpedgiz*

*Spinuzzi, Clay (2003): Tracing Genres through Organizations: A Sociocultural Approach to Information Design (Acting with Technology). The MIT Press*

*Svanaes, Dag (2000): Understanding Interactivity: Steps to a Phenomenology of Human-Computer Interaction. Trondheim, Norway, Norges Teknisk-Naturvitenskapelige Universitet (NTNU)*

*Vasilyuk, Fyodor (1992): The Psychology of Experiencing: The Resolution of Life's Critical Situations (originally published in Russian in 1984). New York University Press*

*Voida, Stephen and Mynatt, Elizabeth D. (2009): It feels better than filing: everyday work experiences in an activity-based computing system. In: Proceedings of ACM CHI 2009 Conference on Human Factors in Computing Systems 2009. pp. 259-268*

*Voida, Stephen, Mynatt, Elizabeth D. and Edwards, W. Keith (2008): Re-framing the desktop interface around the activities of knowledge work. In: Cousins, Steve B. and Beaudouin-Lafon, Michel (eds.) Proceedings of the 21st Annual ACM Symposium on User Interface Software and Technology October 19-22, 2008, Monterey, CA, USA. pp. 211-220*

*Vygotsky, Lev (1978): Mind in Society: The Development of Higher Psychological Processes. Harvard University Press*

*Wertsch, James V. (1981): The concept of activity in Soviet psychology: An introduction. In: Wertsch, James V. (ed.). "The Concept of Activity in Soviet Psychology". M.E. Sharpe*

*Winograd, Terry and Flores, Fernando (1986): Understanding Computers and Cognition. Norwood, NJ, Intellect*

## About the author

### Victor Kaptelinin

*© Victor Kaptelinin*

Has also published under the name of:
"V. Kaptelinin"

Victor Kaptelinin is a Professor at the Department of Information Science and Media Studies, University of Bergen, Norway, and the Department of Informatics, Umeå University, Sweden. He has held teaching and/or research positions at the Psychological Institute of Russian Academy of Education, Moscow Lomonosov University, and University of California in San Diego, USA. His main research interests are in interaction design, activity theory, and educational use of information technologies.

## Your notes and thoughts on chapter 16

Record your notes and thoughts on this chapter. If you want to share these thoughts with others online, go to the bottom of the page at:
http://www.interaction-design.org/encyclopedia/activity_theory.html

**NOTES:**

# Chapter 17

# Disruptive Innovation

### *by Clayton M. Christensen.*

A disruptive technology or disruptive innovation is an innovation that helps create a new market and value network, and eventually goes on to disrupt an existing market and value network. The term is used in business and technology literature to describe innovations that improve a product or service in ways that the market does not expect. Although the term *disruptive technology* is widely used, *disruptive innovation* seems a more appropriate term in many contexts since few technologies are intrinsically disruptive; rather, it is the business model that the technology enables that creates the disruptive impact.

## 17.1 INTRODUCTION

*How can I beat my most powerful competitor? How can I know in advance of the battle whether I'm going to be able to beat the competition? Why has disruption proven to be such a consistently effective strategy for causing strong incumbent* *competitors to flee from their entrant attackers, rather than fight them? How can I shape a business idea into a disruptive strategy?*

What if you could predict the winners in a race for innovative growth? What if you could choose your competitive battles knowing you would win nearly every time? What if you knew in advance which growth strategies would succeed, and which would fail?

Managers have long sought ways to predict the outcome of competitive fights. Some look at the attributes of the companies involved: Larger companies with more resources to throw at a problem will beat the smaller competitors. It's interesting how often the CEOs of large, resource-rich companies base their strategies upon this theory, despite repeated evidence that the level of resources committed often bears little relationship to the outcome.

Others consider the attributes of the change: When innovations are incremental, the established, leading firms in an industry are likely to reinforce their dominance; however, compared with entrants, they will be conservative and ineffective in exploiting breakthrough innovation.[1]

1. We mentioned in the introduction that in early stages of theory building, the best that scholars can do is suggest categories that are defined by the attributes of the phenomena. Such studies are important stepping stones in the path of progress. One such important book is Richard Foster, *Innovation: The Attacker's Advantage* (Foster 1986). Another study predicted that the leaders will fail when an innovation entails development of completely new technological competencies. See Michael L. Tushman and Philip Anderson, "Technological Discontinuities and Organizational Environments," *Administrative Science Quarterly* 31 (1986). The research of MIT Professor James M. Utterback and his colleagues on dominant designs has been particularly instrumental in moving this body of theory toward circumstance-based categorization. See, for example, James M. Utterback and William J. Abernathy, "A Dynamic Model of Process and Product Innovation" *Omega* 33, no. 6 (1975): 639–656; and Clayton M. Christensen, Fernando F. Suarez, and James M. Utterback, "Strategies for Survival in Fast-Changing Industries," *Management Science* 44, no.12 (2001): s207-s2202. Demanding customers are those customers who are willing to pay for increases on some dimension of performance—faster speeds, smaller sizes, better reliability, and so on. Less-demanding or undemanding customers are those customers who would rather make a different trade-off, accepting less performance (slower speeds, larger sizes, less reliability, and so on) in exchange for commensurately lower prices. We depict these trajectories as straight lines because empirically, when charted on semi-long graph paper, they in fact are straight, suggesting that our ability to utilize improvement increases at an exponential pace—though a pace that is shallower than the trajectory of technological progress.

Our ongoing study of innovation suggests another way to understand when incumbents will win, and when the entrants are likely to beat them. *The Innovator's Dilemma (Christensen 1997)* identified two distinct categories—sustaining and disruptive—based on the *circumstances* of innovation. In *sustaining situations*—when the race entails making better products that can be sold for more money to attractive customers—we found that incumbents almost always prevail. In *disruptive circumstances*—when the challenge is to commercialize a simpler, more convenient product that sells for less money and appeals to a new or unattractive customer set—the entrants are likely to beat the incumbents. This is the phenomenonthat so frequently defeats successful companies. It implies, of course, that the best way for upstarts to attack established competitors is to disrupt them.

Few technologies or business ideas are intrinsically sustaining or disruptive in character. Rather, their disruptive impact must be molded into strategy as managers shape the idea into a plan and then implement it. Successful new-growth builders know—either intuitively or explicitly—that disruptive strategies greatly increase the odds of competitive success.

This chapter's purpose is to review the disruptive innovation model from the perspective of both the disruptee *and* the disruptor in order to help growth builders shape their strategies so that they pick disruptive fights they can win. Because disruption happens whether we want it or not, this chapter should also help established companies capture disruptive growth, instead of getting killed by it.

## 17.2 THE DISRUPTIVE INNOVATION MODEL

*The Innovator's Dilemma* (Christensen 1997) identified three critical elements of disruption, as depicted in Figure 17.1. First, in every market there is a rate of improvement that customers can utilize or absorb, represented by the dotted line sloping gently upward across the chart. For example, the automobile companies keep giving us new and improved engines, but we can't utilize all the performance that they make available under the hood. Factors such as traffic jams, speed limits, and safety concerns constrain how much performance we can use.

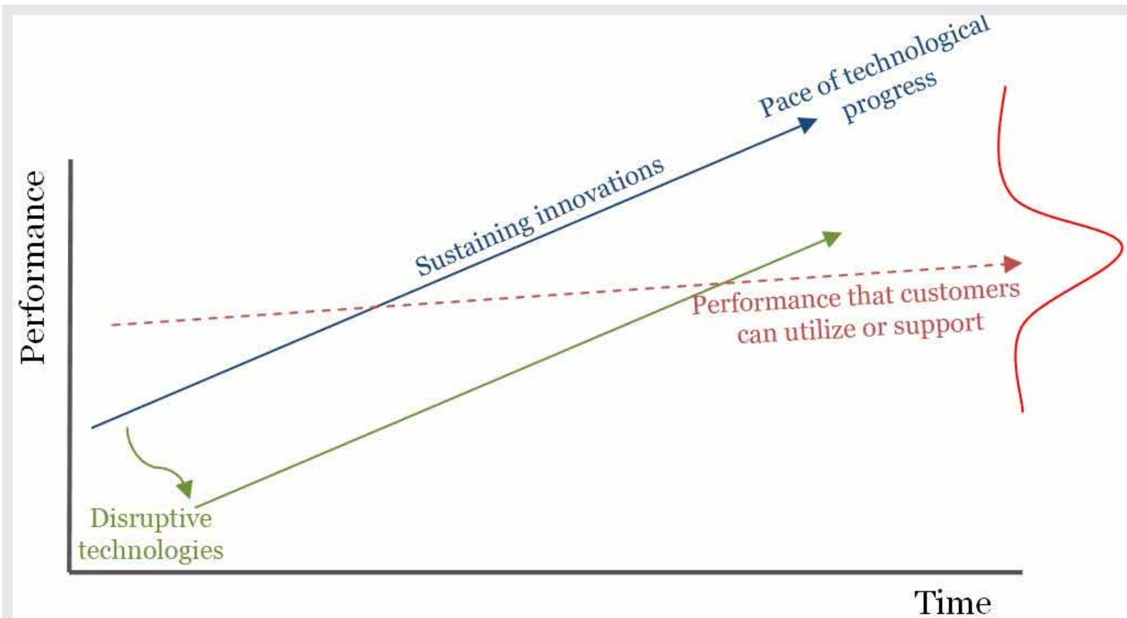



Figure 17.1: The Disruptive Innovation Model.



To simplify the chart, we depict customers' ability to utilize improvement as a single line. In reality, there is a distribution of customers around this median: There are many such lines, or tiers, in a market—a range indicated by the distribution curve at the right. Customers in the highest or most demanding tiers may never be satisfied with the best that is available, and those in the lowest or least demanding tiers can be over-satisfied with very little. But on average, this dotted line represents technology that is "good enough" to serve existing mainstream customers' needs.

Second, in every market there is a distinctly different trajectory of improvement that innovating companies provide as they introduce new and improved products. The more steeply sloping solid lines in Figure 17.1 suggest that this pace of technological progress almost always outstrips the ability of customers in any given tier of the market to use it. Thus, a company whose products are squarely positioned on mainstream customers' current needs will probably overshoot what those same customers are able to utilize in the future. This happens because companies keep striving to make better products that they can sell for higher profit margins to not-yet-satisfied customers in more demanding tiers of the market.

To visualize this, think back to 1983 when people first started using personal computers for word processing. Typists often had to stop their fingers to let the Intel 286 chip inside catch up. As depicted at the left side of Figure 17.1, the technology was not good enough. But today's processors offer much more speed than mainstream customers can use—although there are still a few unsatisfied customers in the most demanding tiers of the market who need even-faster chips.

The third critical element of the model is the distinction between sustaining and disruptive innovation. A *sustaining innovation* targets demanding, high-end customers with better performance than what was previously available. Some sustaining innovations are the incremental year-by-year improvements that all good companies grind out. Other sustaining innovations are breakthrough, leapfrog-beyond-the-competition products. It doesn't matter how technologically difficult the innovation is, however: The established competitors almost always win the battles of sustaining technology. Because this strategy entails making a better product that they can sell for higher profit margins to their best customers, the established competitors have powerful motivations to fight sustaining battles. And they have the resources to win.

*Disruptive innovations,* in contrast, don't attempt to bring better products to established customers in existing markets. Rather, they disrupt and redefine that trajectory by introducing products and services that are not as good as currently available products. But disruptive technologies offer other benefits—typically, they are simpler, more convenient, and less expensive products that appeal to new or less-demanding customers.[2]

Once the disruptive product gains a foothold in new or low-end markets, the improvement cycle begins. And because the pace of technological progress outstrips customers' abilities to use it, the previously not-good-enough technology eventually improves enough to intersect with the needs of more demanding customers. When that happens, the disruptors are on a path that will ultimately crush the incumbents. This distinction is important for innovators seeking to create new-growth businesses. Whereas the current leaders of the industry almost always triumph in battles of sustaining innovation, the odds at disruptive innovation heavily favor entrant companies.[3]

Disruption has a paralyzing effect on industry leaders. With resource allocation processes designed and perfected to support sustaining innovations, they are

2. After watching students and managers read, interpret, and talk about this distinction between sustaining and disruptive technologies, we have observed a stunningly common human tendency to take a new concept, new data, or new way of thinking and morph it so that it fits one's existing mental models. Hence, many people have equated our use of the term *sustaining innovation* with their preexisting frame of "incremental" innovation, and they have equated the term *disruptive technology* with the words *radical, breakthrough, out-of-the-box,* or *different*. They then conclude that disruptive ideas (as they define the term) are good and merit investment. We regret that this happens, because our findings relate to a very specific definition of disruptiveness, as stated in our text here.

3. *The Innovator's Dilemma* notes that the only times that established companies succeeded in staying atop their industries when confronted by disruptive technologies were when the established firms created a completely separate organization and gave it an unfettered charter to build a completely new business with a completely new business model. Hence, IBM was able to remain atop its industry when minicomputers disrupted mainframes because it competed in the minicomputer market with a different business unit. And when the personal computer emerged, IBM addressed that disruption by creating an autonomous business unit in Florida. Hewlett-Packard remained the leader in printers for personal computing because it created a division to make and sell ink-jet printers that was completely independent from its printer division in Boise, which made and sold laser jet printers. Since publication of *The Innovator's Dilemma,* a number of companies that were faced with disruption have succeeded in becoming leaders in the wave of disruption coming at them by setting up separate organizational units to address the disruption. Charles Schwab became the leading online broker; Teradyne, the maker of semiconductor test equipment, became the leader in PC-based testers; and Intel introduced its Celeron chip, which reclaimed the low end of the microprocessor market. We hope that as more established companies learn to address disruptions through independent business units when faced with disruptive opportunities, the odds that historically were overwhelmingly favorable to entrant firms and their venture capital backers will become more favorable to established leaders who seek to create new-growth opportunities.

constitutionally unable to respond. They are always motivated to go up-market, and almost never motivated to defend the new or low-end markets that the disruptors find attractive. We call this phenomenon *asymmetric motivation*. It is the core of the innovator's dilemma, and the beginning of the innovator's solution.

### 17.2.1 Disruption at Work: How Minimills Upended Integrated Steel Companies

The disruption of integrated steel mills by minimills, which is reviewed briefly in *The Innovator's Dilemma (Christensen 1997)*, offers a classic example of why established leaders are so much easier to beat if the idea for a new product or business is shaped into a disruption.

Historically, most of the world's steel has come from massive integrated mills that do everything from reacting iron ore, coke, and limestone in blast furnaces to rolling finished products at the other end. It costs about $8 billion to build a huge new integrated mill today. Minimills, in contrast, melt scrap steel in electric arc furnaces—cylinders that are approximately twenty meters in diameter and ten meters tall. Because they can produce molten steel cost-effectively in such a small chamber, minimills don't need the massive-scale rolling and finishing operations that are required to handle the output of efficient blast furnaces—which is why they are called *minimills*. Most important, though, minimills' straightforward technology can make steel of any given quality for 20 percent lower cost than an integrated mill.

Steel is a commodity. You would think that every integrated steel company in the world would have aggressively adopted the straightforward, lower-cost minimill technology. Yet as of 2000 not a single integrated steel company had successfully invested in a minimill, even as the minimills had grown to account for nearly half of North America's steel production and a significant share of other markets as well.[4]

4. An exception to this statement is found in Japan, where a couple of integrated mills have subsequently acquired existing minimill companies.

We can explain why something that makes so much sense has been so difficult for the integrated mills. Minimills first became technologically viable in the mid-1960s. Because they melt scrap of uncertain and varying chemistry in their electric arc furnaces, the quality of the steel that minimills initially could produce was poor. In fact, the only market that would accept the output of minimills was the concrete reinforcing bar (rebar) market. The specifications for rebar are loose, so this was an ideal market for products of low and variable quality.

As the minimills attacked the rebar market, the integrated mills were happy to be rid of that dog-eat-dog commodity business. Because of the differences in their cost structures and the opportunities for investment that they each faced, the rebar market looked very different to the disruptee and the disruptor. For integrated producers, gross profit margins on rebar often hovered near 7 percent, and the entire product category accounted for only 4 percent of the industry's tonnage. It was the least attractive of any tier of the market in which they might invest to grow. So as the minimills established a foothold in the rebar market, the integrated mills reconfigured their rebar lines to make more profitable products.

In contrast, with a 20 percent cost advantage, the minimills enjoyed attractive profits in competition against the integrated mills for rebar—until 1979, when the minimills finally succeeded in driving the last integrated mill out of the rebar market. Historical pricing statistics show that the price of rebar then collapsed by 20 percent. As long as the minimills could compete against higher-cost integrated mills, the game was profitable for them. But as soon as low-cost minimill was pitted against low-cost minimill in a commodity market, the reward for victory was that none of them could earn attractive profits in rebar.[5] Worse, as

5. The economists' simple notion that price is determined at the intersection of supply and demand curves explains this phenomenon. Price gravitates to the cash cost of the marginal, or highest-cost, producer whose capacity is required for supply to meet the quantity demanded. When the marginal producers were high-cost integrated mills, minimills could make money in rebar. When the marginal, highest-cost producers were minimills, then the price of rebar collapsed. The same mechanism destroyed the temporary profitability to the minimills of each subsequent tier of the market, as described in the text that follows.

they all sought profitability by becoming more efficient producers, they discovered that cost reductions meant survival, but not profitability, in a commodity such as rebar.[6]

Soon, however, the minimills looked up-market, and what they saw there spelled relief. If they could just figure out how to make bigger and better steel—shapes like angle iron and thicker bars and rods—they could roll *tons* of money, because in that tier of the market, as suggested in Figure 17.2, the integrated mills were earning gross margins of about 12 percent—nearly double the margins that they had been able to earn in rebar. That market was also twiceas big as the rebar segment, accounting for about 8 percent of industry tonnage. As the minimills figured out how to make bigger and better steel and attacked that tier of the market, the integrated mills were almost relieved to be rid of the bar and rod business as well. It was a dog-eat-dog commodity compared with their higher-margin products, whereas for the minimills, it was an attractive opportunity compared with their lower-margin rebar. So as the minimills expanded their capacity to make angle iron and thicker bars and rods, the integrated mills shut their lines down or reconfigured them to make more profitable products. With a 20 percent cost advantage, the minimills enjoyed significant profits in competition against the integrated mills until 1984, when they finally succeeded in driving the last integrated mill out of the bar and rod market. Once again, the minimills reaped their reward: With low-cost minimill pitted against low-cost minimill, the price of bar and rod collapsed by 20 percent, and they could no longer earn attractive profits. What could they do?

6. That cost reduction rarely creates competitive advantage is argued persuasively in Michael Porter, "What Is Strategy?" *Harvard Business Review*, November–December 1996, 61–78.

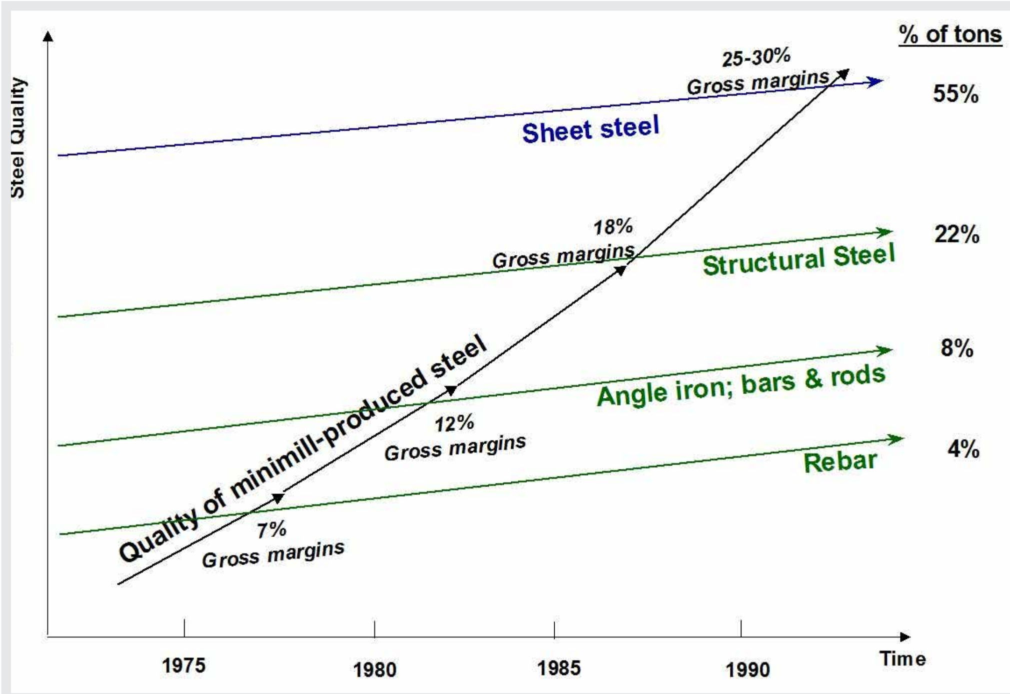



Figure 17.2: The Disruptive Attack of the Steel Minimills.



Continued up-market movement into structural beams appeared to be the next obvious answer. Gross margins in that sector were a whopping 18 percent, and the market was three times as large as the bar and rod business. Most industry technologists thought minimills would be unable to roll structural beams. Many of the properties required to meet the specifications for steel used in building and bridge construction were imparted to the steel in the rolling processes of big integrated mills, and you just couldn't get those properties in minimills' abbreviated facilities. What the technical experts didn't count on, however, was how desperately motivated the minimills would be to solve that problem, because it was the only way they could make attractive money. Minimills achieved extraordinarily clever innovations as they stretched from angle iron to I-beams—things such as Chaparral Steel's dog-bone mold in its continuous caster, which no one had imagined could be done. Although you could never have predicted what the technical solution would be, you *could* predict with perfect certainty that the minimills were powerfully motivated to figure it out. Necessity remains the mother of invention.

At the beginning of their invasion into structural beams, the biggest that the minimills could roll were little six-inch beams of the sort that under-gird mobile homes. They attacked the low end of the structural beam market, and again the integrated mills were almost relieved to be rid of it. It was a dog-eat-dog commodity compared with their other higher-margin products where focused investment might bring more attractive volume. To the minimills, in contrast, it was an attractive product compared with the margins they were earning on rebar and angle iron. So as the minimills expanded their capacity to roll structural beams, the integrated mills shut their structural beam mills down in order to focus on more profitable sheet steel products. With a 20 percent cost advantage, the minimills enjoyed significant profits as long as they could compete against the integrated mills. Then in the mid-1990s, when they finally succeeded in driving the last integrated mill out of the structural beam market, pricing again collapsed. Once again, the reward for victory was the end of profit.

The sequence repeated itself when the leading minimill, Nucor, attacked the sheet steel business. Its market capitalization now dwarfs that of the largest integrated steel company, US Steel. Bethlehem Steel is bankrupt at the time of this writing.

This is not a history of bungled steel company management. It is a story of rational managers facing the innovator's dilemma: Should we invest to protect the least profitable end of our business, so that we can retain our least loyal, most price-sensitive customers? Or should we invest to strengthen our position in the most profitable tiers of our business, with customers who reward us with premium prices for better products?

The executives who confront this dilemma come in all varieties: timid, feisty, analytical, and action-driven. In an unstructured world their actions might be unpredictable. But as large industry incumbents, they encounter powerful and predictable forces that motivate them to flee rather than fight when attacked from below. That is why shaping a business idea into a disruption is an effective strategy for beating an established competitor. Disruption works because it is *much* easier to beat competitors when they are motivated to flee rather than fight.

The forces that propel well-managed companies up-market are *always* at work, in every company in every industry. Whether or not entrant firms have disrupted the established leaders yet, the forces are at work, leading predictably in one direction. It is not just a phenomenon of "technology companies" such as those involved in microelectronics, software, photonics, or biochemistry. Indeed, when we use the term *technology* in this chapter, it means the process that *any* company uses to convert inputs of labor, materials, capital, energy, and information into outputs of greater value. For the purpose of predictably creating growth, treating "high tech" as different from "low tech" is not the right way to categorize the world. Every company has technology, and each is subject to these fundamental forces.

### 17.2.2 The Role of Sustaining Innovation in Generating Growth

We must emphasize that we do *not* argue against the aggressive pursuit of sustaining innovation. Several other insightful books offer management techniques to help companies excel in sustaining innovations—and their contribution is important.[7] Almost always a host of similar companies enters an industry in its early years, and getting ahead of that crowd—moving up the sustaining-innovation trajectory more decisively than the others—is critical to the successful *exploitation* of the disruptive opportunity. But this is the source of the dilemma: Sustaining innovations are so important and attractive, relative to disruptive ones, that the very best sustaining companies systematically ignore disruptive threats and opportunities until the game is over.

7. We recommend in particular Steven C. Wheelwright and Kim B. Clark, *Revolutionizing New Product Development* (New York: The Free Press, 1992); Stefan Thomke, *Experimentation Matters: Unlocking the Potential of New Technologies for Innovation* (Boston: Harvard Business School Press, 2003); Stefan Thomke and Eric von Hippel, Customers as Innovators: A New Way to Create Value" *Harvard Business Review*, 80 No. 4 (April 2002): 74-81; and Eric von Hippel, *The Sources of Innovation*. (New York, Oxford University Press, 1988).

Sustaining innovation essentially entails making a better mousetrap. Starting a new company with a sustaining innovation isn't necessarily a bad idea: Focused companies sometimes can develop new products more rapidly than larger firms because of the conflicts and distractions that broad scope often creates. The theory of disruption suggests, however, that once they have developed and established the viability of their superior product, entrepreneurs who have entered on a sustaining trajectory should turn around and sell out to one of the industry leaders behind them. If executed successfully, getting ahead of the leaders on the sustaining curve and then selling out quickly can be a straightforward way to make an attractive financial return. This is common practice in the health care industry, and was the well-chronicled mechanism by which Cisco Systems "outsourced" (and financed with equity capital, rather than expense money) much of its sustaining-product development in the 1990s.

A sustaining-technology strategy is *not* a viable way to build *new*-growth businesses, however. If you create and attempt to sell a better product into an established market to capture established competitors' best customers, the competitors will be motivated to fight rather than to flee.[8] This advice holds even when the entrant is a huge corporation with ostensibly deeper pockets than the incumbent.

8. This model explains quite clearly why the major airline companies in the United States are so chronically unprofitable. Southwest Airlines entered as a new-market disruptor, competing within Texas for customers who otherwise would not have flown at all, but would have used automobiles and buses. The airline has grown carefully into nonmajor airports, staying away from head-on competition against the majors. It is the low-end disruptors to this industry—airlines with names such as JetBlue, AirTran, People Express, Florida Air, Reno Air, Midway, Spirit, Presidential, and many others—that create the chronic unprofitability.
When leaders in most other industries get attacked by low-end disruptors, they can run away up-market and remain profitable (and often improve profitability) for some time. The integrated steel companies fled up-market away from the minimills. The full-service department stores fled up-market into clothing, home furnishings, and cosmetics when the discount department stores attacked branded hard goods such as hardware, paint, toys, sporting goods, and kitchen utensils at the low-margin end of the merchandise mix. Today, the discount department stores such as Target and Wal-Mart are fleeing up-market into clothing, home furnishings, and cosmetics as hard goods discounters such as Circuit City, Toys 'R Us, Staples, Home Depot, and Kitchens Etc. attack the low end; and so on.
The problem in airlines is that the majors cannot flee up-market. Their high fixed-cost structure makes it impossible to abandon the low end. Hence, low-end disruptors easily enter and attack; once one of them gets big enough, however, the major airlines declare that enough is enough, and they turn around and fight. This is why no low-end disruptor to date has survived for longer than a few years. But because low-end disruption by new companies is so easy to start, the majors can never raise low-end pricing up to levels of attractive profitability.

For example, electronic cash registers were a radical but sustaining innovation relative to electromechanical cash registers, whose market was dominated by National Cash Register (NCR). NCR *totally* missed the advent of the new technology in the 1970s—so badly, in fact, that NCR's product sales literally went to zero. Electronic registers were so superior that there was no reason to buy an electromechanical product except as an antique. Yet NCR survived on service revenues for over a year, and when it finally introduced its own electronic cash register, its extensive sales organization quickly captured the same share of the market as the company had enjoyed in the electromechanical realm.[9] The attempts that IBM and Kodak made in the 1970s and 1980s to beat Xerox in the high-speed photocopier business are another example. These companies were *far* bigger, and yet they failed to outmuscle Xerox in a sustaining-technology competition. The firm that beat Xerox was Canon—and that victory started with a disruptive tabletop copier strategy.

Similarly, corporate giants RCA, General Electric, and AT&T failed to outmuscle IBM on the sustaining-technology trajectory in mainframe computers. Despite the massive resources they threw at IBM, they couldn't make a dent in IBM's position. In the end, it was the disruptive personal computer makers, not the major corporations who picked a direct, sustaining-innovation fight, that bested IBM in computers. Airbus entered the commercial airframe industry head-on against Boeing, but doing so required massive subsidies from European governments. In the future, the most profitable growth in the airframe industry will probably come from firms with disruptive strategies, such as Embraer and Bombardier's Canadair, whose regional jets are aggressively stretching up-market from below.[10]

9. This history is recounted in a marvelous paper by Richard S. Rosenbloom, "From Gears to Chips: The Transformation of NCR and Harris in the Digital Era," working paper, Harvard Business School Business History Seminar, Boston, 1988.

10. We would be foolish to claim that it is impossible to create new-growth companies with a sustaining, leap-beyond-the-competition strategy. It is more accurate to say that the odds of success are very, very low. But some sustaining entrants have succeeded. For example, EMC Corporation took the high-end

### 17.2.3 Disruption Is a Relative Term

An idea that is disruptive to one business may be sustaining to another. Given the stark odds that favor the incumbents in the sustaining race but entrants in disruptive ones, we recommend a strict rule: If your idea for a product or business appears disruptive to some established companies but might represent a *sustaining* improvement for others, then you should go back to the drawing board. You need to define an opportunity that is disruptive relative to *all* the established players in the targeted market space, or you should not invest in the idea. If it is a sustaining innovation relative to the business model of a significant incumbent, you are picking a fight you are very unlikely to win.

Take the Internet, for example. Throughout the late 1990s, investors poured billions into Internet-based companies, convinced of their "disruptive" potential. An important reason why many of them failed was that the Internet was a sustaining innovation relative to the business models of a host of companies. Prior to the advent of the Internet, Dell Computer, for example, sold computers directly to customers by mail and over the telephone. This business was already a low-end disruptor, moving up its trajectory. Dell's banks of telephone salespeople had to be highly trained in order to walk their customers through the various configurations of components that were and were not feasible. They then manually entered the information into Dell's order fulfillment systems.

data storage business away from IBM in the 1990s with a different product architecture than IBM's. But as best we can tell, EMC's products were *better* than IBM's in the very applications that IBM served. Hewlett-Packard's laser jet printer business was a sustaining technology relative to the dot-matrix printer, a market dominated by Epson. Yet Epson missed it. The jet engine was a radical but sustaining innovation relative to the piston aircraft engine. Two of the piston engine manufacturers, Rolls-Royce and Pratt & Whitney, navigated the transition to jets successfully. Others, such as Ford, did not. General Electric was an entrant in the jet revolution, and became very successful. These are anomalies that the theory of disruption cannot explain. Although our bias is to assume that most managers most of the time are on top of their businesses and manage them in competent ways, it is also true that sometimes managers simply fall asleep at the switch.

For Dell, the Internet was a sustaining technology. It made Dell's core business processes work better, and it helped Dell make more money in the way it was structured to make money. But the identical strategy of selling directly to customers over the Internet was *very* disruptive relative to Compaq's business model, because that company's cost structure and business processes were targeted at in-store retail distribution.

The theory of disruption would conclude that if Dell (and Gateway) had not existed, then start-up Internet-based computer retailers might have succeeded in disrupting competitors such as Compaq. But because the Internet was sustaining to powerful incumbents, entrant Internet computer retailers have not prospered.

### 17.2.4 A Disruptive Business Model Is a Valuable Corporate Asset

A disruptive business model that can generate attractive profits at the discount prices required to win business at the low end is an extraordinarily valuable growth asset. When its executives carry the business model up-market to make higher-performance products that sell at higher price points, much of the increment in pricing falls to the bottom line—and it continues to fall there as long as the disruptor can keep moving up, competing at the margin against the higher-cost disruptee. When a company tries to take a higher-cost business model down-market to sell products at lower price points, almost none of the incremental revenue will fall to its bottom line. It gets absorbed into overheads. This is why established firms that hope to capture the growth created by disruption need to do so from within an autonomous business with a cost structure that offers as much headroom as possible for subsequent profitable migration up-market.

Moving up the trajectory into successively higher-margin tiers of the market and shedding less-profitable products at the low end is something that all good managers must do in order to keep their margins strong and their stock price healthy. Standing still is not an option, because firms that stop moving up find themselves in a rebar-esque situation, slugging it out with hard-to-differentiate products against competitors whose costs are comparable[11].

This ultimately means that in doing what they must do, every company prepares the way for its own disruption. This is the innovator's dilemma. But it also is the beginning of the innovator's solution. It does not guarantee success, but it sure helps: *The Innovator's Dilemma* (Christensen 1997) showed that following a strategy of disruption increased the odds of creating a successful growth business from 6 to 37 percent.[12] Because the established company's course of action is mandated so clearly, it is also clear what executives who seek to create new-growth businesses should do: Target products and markets that the established companies are motivated to ignore or run away from. Many of the most profitable growth trajectories in history have been initiated by disruptive innovations.

## 17.3 TWO TYPES OF DISRUPTION

For the sake of simplicity, *The Innovator's Dilemma* (Christensen 1997) presented the disruptive innovation diagram in only two dimensions. In reality, there are two different types of disruptions, which can best be visualized by adding a third axis to the disruption diagram, as shown in Figure 17.3. The vertical and horizon-

11. This partially explains, for example, why Dell Computer has been such a successful disruptor—because it has raced up-market in order to compete against higher-cost makers of workstations and servers such as Sun Microsystems. Gateway, in contrast, has not prospered to the same extent even though it had a similar initial business model, because it has not moved up-market as aggressively and is stuck with undifferentiable costs selling undifferentiable computers. We believe that this insight represents a useful addendum to Professor Michael Porter's initial notion that there are two viable types of strategy—differentiation and low cost (Michael Porter, *Competitive Strategy*. New York: The Free Press, 1980). The research of disruption adds a dynamic dimension to Porter's work. Essentially, a low-cost strategy yields attractive profitability only until the higher-cost competitors have been driven from a tier in the market. Then, the low-cost competitor needs to move up so that it can compete once again against higher-cost opponents. Without the ability to move up, a low-cost strategy becomes an equal-cost strategy.

12. See Clayton M. Christensen, *The Innovator's Dilemma* (Boston: Harvard Business School Press, 1997), 130.

tal axes are as before: the performance of the product on the vertical axis, with time plotted on the horizontal dimension. The third axis represents new customers and new contexts for consumption.

Our original dimensions—time and performance—define a particular market application in which customers purchase and use a product or service. In geometric terms, this application and set of customers reside in a plane of competition and consumption, which *The Innovator's Dilemma* called a *value network*. A value network is the context within which a firm establishes a cost structure and operating processes and works with suppliers and channel partners in order to respond profitably to the common needs of a class of customers. Within a value network, each firm's competitive strategy, and particularly its cost structure and its choices of markets and customers to serve, determines its perceptions of the economic value of an innovation. These perceptions, in turn, shape the rewards and threats that firms expect to experience through disruptive versus sustaining innovations.[13]

The third dimension that extends toward us in the diagram represents new contexts of consumption and competition, which are new value networks. These constitute either new customers who previously lacked the money or skills to buy

13. The concept of value networks was introduced in Clayton M. Christensen, "Value Networks and the Impetus to Innovate," chapter 2 in *The Innovator's Dilemma*. Professor Richard S. Rosenbloom of the Harvard Business School originally identified the existence of value networks when he advised Christensen's early research. In many ways, the situation in a value network corresponds to a "Nash equilibrium," developed by Nobel Laureate John Nash (who became even more renowned through the movie *A Beautiful Mind*). In a Nash equilibrium, given Company A's understanding of the optimal, self-interested (maximum-profit) strategy of each of the other companies in the system, Company A cannot see any better strategy for itself than the one it presently is pursuing. The same holds true for all other companies in the system. Hence, none of the companies is motivated to change course, and the entire system therefore is relatively inert to change. Insofar as the companies within a value network are in a Nash equilibrium, it creates a drag that constrains how fast customers can begin utilizing new innovations. This application of Nash equilibriums to the uptake of innovations was recently introduced in Bhaskar Chakravorti, *The Slow Pace of Fast Change* (Boston: Harvard Business School Press, 2003). Although Chakravorti did not make the linkage himself, his concept is a good way to visualize two things about the disruptive innovation model. It explains why the pace of technological progress outstrips the abilities of customers to utilize the progress. It also explains why competing against nonconsumption, creating a completely new value network, is often in the long run an easier way to attack an established market.

and use the product, or different situations in which a product can be used—enabled by improvements in simplicity, portability, and product cost. For each of these new value networks, a vertical axis can be drawn representing a product's performance as it is defined in that context (which is a different measure from what is valued in the original value network).

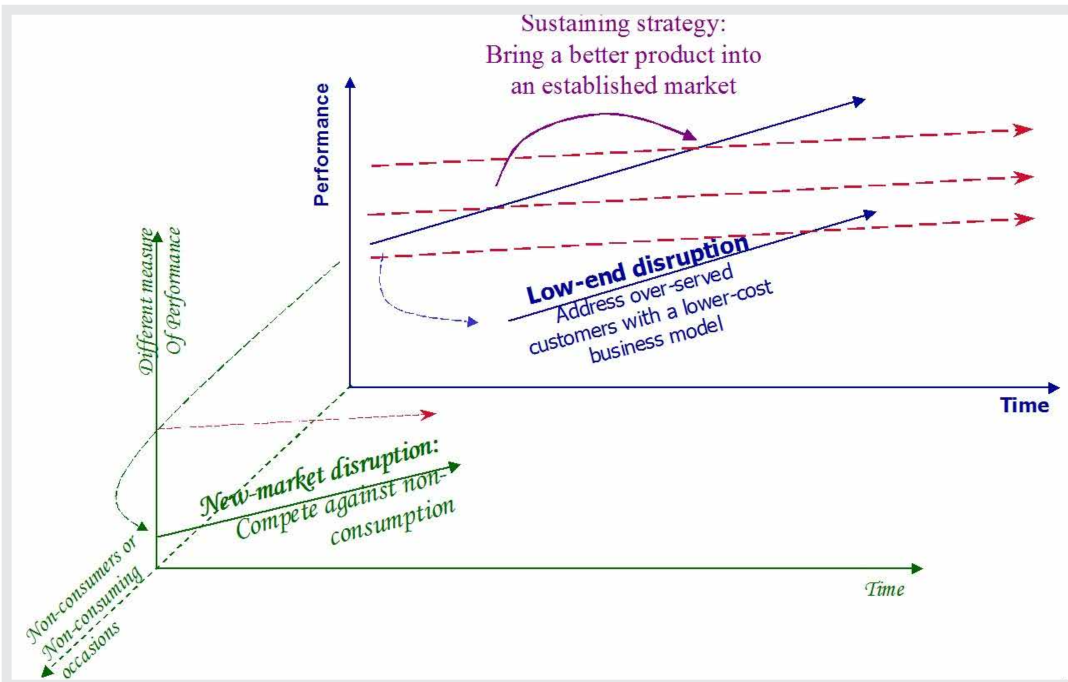



Figure 17.3: Two Types of Disruptive Innovations.



Different value networks can emerge at differing distances from the original one along the third dimension of the disruption diagram. In the following discussion, we will refer to disruptions that create a new value network on the third axis as *new-market disruptions*. In contrast, *low-end disruptions* are those that attack the least-profitable and most overserved customers at the low end of the original value network.

### 17.3.1 New-Market Disruptions

We say that new-market disruptions compete with "nonconsumption" because new-market disruptive products are so much more affordable to own and simpler to use that they enable a whole new population of people to begin owning and using the product, and to do so in a more convenient setting. The personal computer and Sony's first battery-powered transistor pocket radio were new-market disruptions, in that their initial customers were new consumers—they had not owned or used the prior generation of products and services. Canon's desktop photocopiers were also a new-market disruption, in that they enabled people to begin conveniently making their own photocopies right in their offices, rather than taking their originals to the corporate high-speed photocopy center where a technician had to run the job for them. When Canon made photocopying so convenient, people ended up making *a lot* more copies. New-market disruptors' challenge is to create a new value network, where it is non-consumption, not the incumbent, that must be overcome.

Although new-market disruptions initially compete against non-consumption in their unique value network, as their performance improves they ultimately become good enough to pull customers out of the original value network into the new one, starting with the least-demanding tier. The disruptive innovation doesn't invade the mainstream market; rather, it pulls customers out of the mainstream market into the new one because these customers find it more convenient to use the new product.

Because new-market disruptions compete against non-consumption, the incumbent leaders feel no pain and little threat until the disruption is in its final stages. In fact, when the disruptors begin pulling customers out of the low end of the original value network, it actually feels good to the leading firms, because as they move up-market in their own world, for a time they are replacing the low-margin revenues that they lose to the disruptors with higher-margin revenues.[14]

14. Some people have concluded on occasion that when the incumbent leader doesn't instantly get killed by a disruption, the forces of disruption somehow have ceased to operate, and that the attackers are being

## 17.3.2 Low-End Disruptions

We call disruptions that take root at the low end of the original or mainstream value network *low-end* disruptions. Disruptions such as steel minimills, discount retailing, and the Korean automakers' entry into the North American market have been pure low-end disruptions in that they did not create new markets—they were simply low-cost business models that grew by picking off the least attractive of the established firms' customers. Although they are different, new-market and low-end disruptions both create the same vexing dilemma for incumbents. New-market disruptions induce incumbents to ignore the attackers, and low-end disruptions motivate the incumbents to flee the attack.

Low-end disruption has occurred several times in retailing.[15] For example, full-service department stores had a business model that enabled them to turn inventories three times per year. They needed to earn 40 percent gross margins to make money within their cost structure. They therefore earned 40 percent three times each year, for a 120 percent annual return on capital invested in inventory (ROCII). In the 1960s, discount retailers such as Wal-Mart and Kmart attacked the low end of the department stores' market—nationally branded hard goods such as paint, hardware, kitchen utensils, toys, and sporting goods—that were so familiar in use that they could sell themselves. Customers in this tier of the market were overserved by department stores, in that they did not need well-trained floor

held at bay. (See, for example, Constantinos Charitou and Constantinos Markides, "Responses to Disruptive Strategic Innovation," *MIT Sloan Management Review*, Winter 2003, 55.) These conclusions reflect a shallow understanding of the phenomenon, because disruption is a process and not an event. The forces are operating all of the time in every industry. In some industries it might take decades for the forces to work their way through an industry. In other instances it might take a few years. But the forces—which really are the pursuit of the profit that comes from competitive advantage—are always at work. Similarly, other writers on occasion have noticed that the leader in an industry actually did not get killed by a disruption, but skillfully caught the wave. They then conclude that the theory of disruption is false. This is erroneous logic as well. When we see an airplane fly, it does not disprove the law of gravity. Gravity continues to exert force on the flying plane—it's just that engineers figured out how to deal with the force. When we see a company succeed at disruption, it is because the management team figured out how to harness the forces to facilitate success.

15. See Clayton M. Christensen and Richard S. Tedlow, "Patterns of Disruption in Retailing," *Harvard Business Review*, January–February 2000, 42–45.

salespeople to help them get what they needed. The discounters' business model enabled them to make money at gross margins of about 23 percent, on average. Their stocking policies and operating processes enabled them to turn inventories more than five times annually, so that they also earned about 120 percent annual ROCII. The discounters did not accept lower levels of profitability—their business model simply earned acceptable profit through a different formula.[16]

It is very hard for established firms *not* to flee from a low-end disruptor. Consider, for example, the choice that executives of full-service department stores had to make when the discount retailers were attacking the branded hard goods at the low end of department stores' merchandise mix. Retailers' critical resource allocation decision is the use of floor or shelf space. One option for department store executives was to allocate more space to even higher-margin cosmetics and high-fashion apparel, where gross margins often exceeded 50 percent. Because their business model turned inventories three times annually, this option promised 150 percent ROCII.

The alternative was to defend the branded hard goods businesses, which the discounters were attacking with prices 20 percent below those of department stores. Competing against the discounters at those levels would send margins plummeting to 20 percent, which, given the three-times inventory turns that were on average inherent in their business model, entailed a ROCII of 60 percent. It thus made perfect sense for the full-service department stores to flee—to get out of the very tiers of the market that the discounters were motivated to enter.[17]

---

16. Ultimately, Wal-Mart was able to create processes that turned assets faster than Kmart. This allowed it to earn higher returns at comparable gross profit margins, giving Wal-Mart a higher sustainable growth rate.

17. The reason it is so much easier for firms in the position of the full-service department stores to flee from the disruption rather than stand to fight it is that in the near term, inventory and asset turns are hard to change. The full-service department stores offered to customers a much broader product selection (more SKUs per category), which inevitably depressed inventory turns. Discounters not only offered a narrower range of products that focused only on the fastest-turning items, but also their physical infrastructure typically put all merchandise on the sales floor. Department stores, in contrast, often had to maintain stockrooms to provide back-up for the limited quantities of any given item that could be placed on their SKU-laden shelves. Hence, when disruptive discounters invaded a tier of their mer-

Many disruptions are hybrids, combining new-market and low-end approaches, as depicted by the continuum of the third axis in Figure 17.3. Southwest Airlines is actually a hybrid disruptor, for example. It initially targeted customers who weren't flying—people who previously had used cars and buses. But Southwest pulled customers out of the low end of the major airlines' value network as well. Charles Schwab is a hybrid disruptor. It stole some customers from full-service brokers with its discounted trading fees, but it also created new markets by enabling people who historically were not equity investors—such as students—to begin owning and trading stocks.[18]

Figure 17.4 shows where some of history's more successful disruptors were positioned along the continuum of new-market to low-end disruption at their inception. The appendix to this chapter offers a brief historical explanation of each of the disruptive products or companies listed on the chart. This is not a complete census of disruptive companies, of course, and their position on the chart is only approximate. However, the array does convey our sense that disruption is a primary wellspring of growth. The prevalence of Japanese companies such as Sony, Nippon Steel, Toyota, Honda, and Canon in the period between 1960 and 1980and the absence of new disruptive companies in the 1990s, for example, explain a lot about why Japan's economy has stagnated. Many of its most influential companies grew dramatically by disrupting others; but the structure of Japan's economic system inhibits the creation of new waves of disruptive growth that might threaten these same companies today.[19]

chandise mix from below, the department stores could not readily drop margins and accelerate turns. Moving up-market where margins still were adequate was always the more feasible and attractive alternative.

18. Low-end disruptions are a direct example of what economist Joseph Schumpeter termed "creative destruction." Low-end disruptions create a step-change cost reduction within an industry—but it is achieved by entrant firms destroying the incumbents. New-market disruption, in contrast, entails a period of substantial creative creation—new consumption—before the destruction of the old occurs

19. For a deeper exploration of the macroeconomic impact of disruption, see Clayton M. Christensen, Stuart L. Hart, and Thomas Craig, "The Great Disruption," *Foreign Affairs* 80, no.2, March–April 2001,

The chart also shows that disruption is an ongoing force that is always at work—meaning that disruptors in one generation become disruptees later. The Ford Model T, for example, created the first massive wave of disruptive growth in automobiles. Toyota, Nissan, and Honda then created the next wave, and Korean automakers Hyundai and Kia have now begun the third. AT&T's wireline long distance business, which disrupted Western Union, is being disrupted by wireless long distance. Plastics makers such as Dow, DuPont, and General Electric continue to disrupt steel, even as their low end is being eaten away by suppliers of blended polyolefin plastics such as Himont.

80-95; and Stuart L. Hart and Clayton M. Christensen, "The Great Leap: Driving Innovation from the Base of the Pyramid," *MIT Sloan Management Review*, Fall 2002, 51–56. The *Foreign Affairs* paper asserts that disruption was the fundamental engine of Japan's economic miracle of the 1960s, 1970s, and 1980s. Like other companies, these disruptors—Sony, Toyota, Nippon Steel, Canon, Seiko, Honda, and others—have soared to the high end, now producing some of the world's highest-quality products in their respective markets. Like the American and European companies that they disrupted, Japan's giants are now stuck at the high end of their markets, where there is no growth. The reason America's economy did not stagnate for an extended period after its leading companies got pinned to the high end was that people could leave those companies, pick up venture capital on the way down, and start new waves of disruptive growth. Japan's economy, in contrast, lacks the labor market mobility and the venture capital infrastructure to enable this. Hence, Japan played the disruptive game once and profited handsomely. But it is stuck. There truly seem to be microeconomic roots to the country's macroeconomic malaise. The *Sloan* paper builds upon the *Foreign Affairs* piece, asserting that today's developing nations are an ideal initial market for many disruptive innovations; and that disruption is a viable economic development policy.

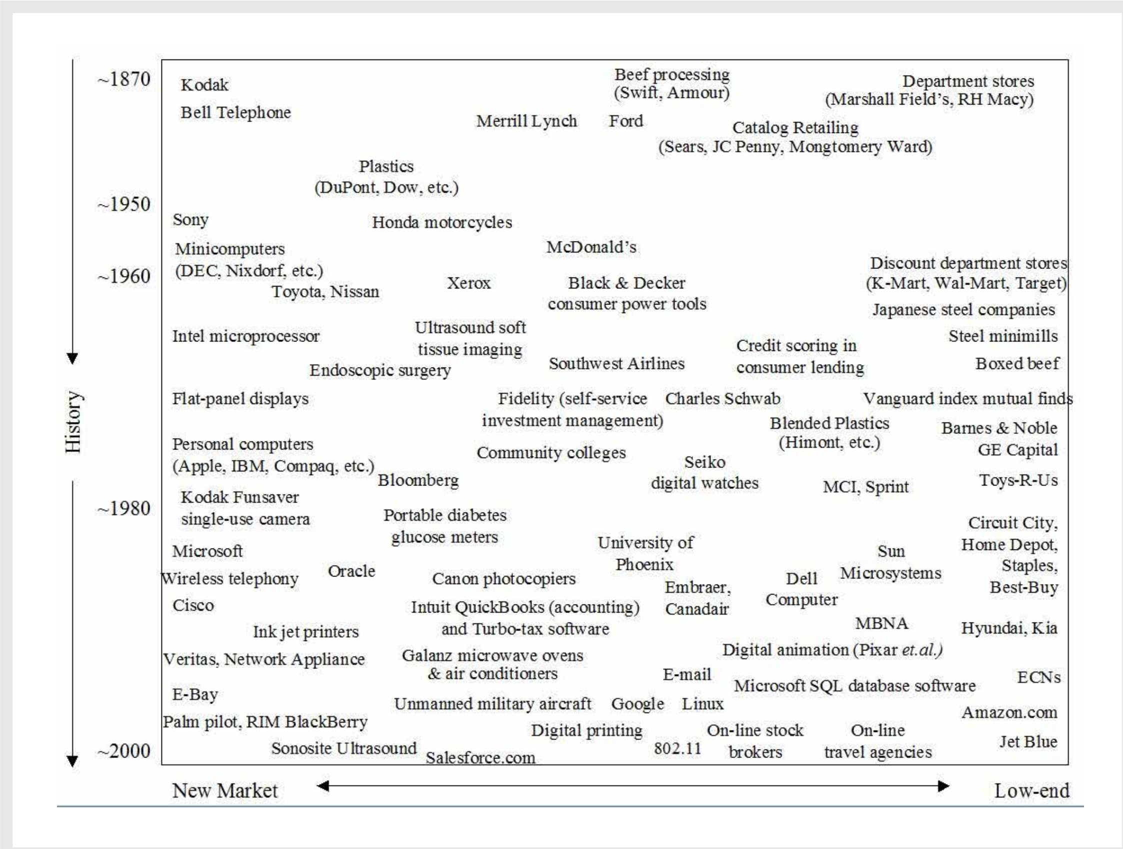



FIGURE 17.4: A Sampling of Companies Whose Origins Were in Disruption.



## 17.4 SHAPING IDEAS TO BECOME DISRUPTIVE: THREE LITMUS TESTS

At the beginning of this chapter, we mentioned that few technologies or product ideas are inherently sustaining or disruptive when they emerge from the innovator's mind. Instead, they go through a process of becoming fleshed out and shaped into a strategic plan in order to win funding. Many—but not all—of the initial ideas that get shaped into sustaining innovations could just as readily be shaped into disruptive business plans with far greater growth potential. The shaping process must be consciously managed, however, and not left in an autopilot mode.

Executives must answer three sets of questions to determine whether an idea has disruptive potential. The first explores whether the idea can become a new-market disruption. For this to happen, at least one and generally both of two conditions must be satisfied:

1. Is there a large population of people who historically have not had the money, equipment, or skill to do this thing for themselves, and as a result have gone without it altogether or have needed to pay someone with more expertise to do it for them?

2. To use the product or service, do customers need to go to an inconvenient, centralized location?

If the technology can be developed so that a large population of less skilled or less affluent people can begin owning and using, in a more convenient context, something that historically was available only to more skilled or more affluent people in a centralized, inconvenient location, then there is potential for shaping the idea into a new-market disruption.

The second set of questions explores the potential for a low-end disruption. This is possible if these two conditions exist:

1. Are there customers at the low end of the market who would be happy to purchase a product with less (but good enough) performance if they could get it at a lower price?

2. Can we create a business model that enables us to earn attractive profits at the discount prices required to win the business of these overserved customers at the low end?

Often, the innovations that enable low-end disruption are improvements in manufacturing, service, or business processes, which enable a company to earn attractive returns on lower gross margins, coupled with processes that turn assets faster.

Once an innovation passes the new-market or low-end test, there is still a third critical consideration, or litmus test, to apply:

1. Is the innovation disruptive to *all* of the significant incumbent firms in the industry? If it appears to be sustaining to one or more significant players in the industry, then the odds will be stacked in that firm's favor, and the entrant is unlikely to win.

If an idea fails the litmus tests, then it cannot be shaped into a disruption. It may have promise as a sustaining technology, but in that case we would expect that it could not constitute the basis of a new-growth business for an entrant company.

For summary, Table 17.1 contrasts the characteristics of the three strategies that firms might pursue in creating new-growth businesses: sustaining innovations, new-market disruptions, and low-end disruptions. It compares the targeted product performance or features, the targeted customers or markets, and the business model implications that each route entails. We hope that managers can use this as a template so that they can categorize and see the implications of different plans that might be presented to them for approval.

Executives can use this categorization and the litmus tests to foresee the competitive consequences of alternative strategies as they shape an idea. To illustrate, we'll examine three questions: whether Xerox could disrupt Hewlett-Packard's ink-jet printing business, how to create growth in air conditioning, and whether online banking had (or has) the disruptive potential to create a new-growth business.

| Dimension | Sustaining Innovations | Low-end disruptions | New Market Disruptions |
|---|---|---|---|
| **Targeted performance of the product or service** | Results in performance improvement **in attributes most valued by the industry's mainstream customers.** These improvements may be incremental or breakthrough in character. | Technology yields products that are good enough along the traditional metrics of performance at the low end of the mainstream market. | Results in lower performance in "traditional" attributes, but **improved performance in new attributes – typically simplicity and convenience.** |
| **Targeted customers or market application** | The **most attractive (i.e., profitable) customers** in the mainstream markets who are willing to pay for improved performance. | Targets **over-served customers** in the low end of the mainstream market. | Targets **non-consumption:** *customers who historically lacked the money or skill to buy and use the product.* |

| | | | |
|---|---|---|---|
| **Impact on the required business model (processes and cost structure)** | Improves or maintains profit margins by exploiting the **existing processes and cost structure,** and making better use of current competitive advantages. | Utilizes a **new operating and / or financial approach** – a different combination of lower gross profit margins and higher asset utilization that can earn attractive returns at the discount prices required to win business at the low end of the market. | Business model must make money at lower price per unit sold, and at unit production volumes that initially will be small emerging market. Gross margin dollars per unit sold will be significantly lower. |

Table 17.1: Distinguishing Characteristics of Sustaining vs. Low-End and New-Market Disruptions.

### 17.4.1 Could Xerox Disrupt Hewlett-Packard?

We don't actually know if Xerox has considered the possibility of creating a new business of the sort we will examine here, and we use the companies' names only to make the example more vivid. We've based this scenario solely on information from public sources. Xerox reportedly has developed outstanding ink-jet printing technology. What can it do with it? It could attempt to leapfrog Hewlett-Packard by making the best ink-jet printer on the market. Even if it could make a better printer, however, Xerox would be fighting a battle of sustaining technology against a company with superior resources and more at stake. HP would win that fight. But could Xerox craft a disruptive strategy for this technology? We'll test the conditions for a low-end strategy first.

To determine whether this strategy is viable, Xerox's managers should test whether customers in the lowest market tiers might be willing to buy a "good enough" printer that is cheaper than prevailing products.[20] At the highest tier of the market, customers seem willing to pay significantly more for a faster printer that produces sharper images. However, consumers in the less-demanding tiers are becoming increasingly indifferent to improvements. It is likely they would be interested in lower-cost alternatives. So the first question gets an affirmative answer.

The next question is whether Xerox could define a business model that could generate attractive returns at the discounted prices required to win business at the low end. The possibilities here don't look good. HP and other printer companies already outsource the fabrication and assembly of components to the lowest-cost sources in the world. HP makes its money selling ink cartridges—whose fabrication also is outsourced to low-cost suppliers. Xerox could enter the market by selling ink cartridges at lower prices, but unless it could define an overhead cost structure and business processes that would allow it to turn assets faster, Xerox could not sustain a strategy of low-end disruption.[21]

20. Our choice of wording in this paragraph is important. When customers cannot differentiate products from each other on any dimension that they can value, then price is often the customer's basis of choice. We would not say, however, that when a consumer buys the lowest-priced alternative, the axis of competition is cost based. The right question to ask is whether customers will be willing to pay higher prices for further improvements in functionality, reliability, or convenience. As long as customers reward improvements with commensurately higher prices, we take it as evidence that the pace of performance improvement has not yet overshot what customers can use. When the marginal utility that customers receive from additional improvements on any of these dimensions approaches zero, then cost is truly the basis of competition.

21. We emphasize the term *product* strategy in this sentence because there certainly seems to be scope for two other low-end disruptive plays in this market. One would be a private-label strategy to disrupt the Hewlett-Packard brand. The other would be a low-cost distribution strategy through an online retailer such as Dell Computer.

This means we'll need to evaluate the potential for a new-market disruption—competing against non-consumption. Is there a large, untapped population of computer owners who don't have the money or skill to buy and use a printer? Probably not. Hewlett-Packard already competed successfully against non-consumption when it launched its easy-to-use, inexpensive ink-jet printers

What about enticing existing printer owners to buy more printers, by enabling consumption in a new, more convenient context? Now, this might be achievable. Documents created on notebook computers are not easy to print. Notebook users have to find a stationary printer and connect to it either over a network or a printer cable, or they must transfer the file via removable media to a computer that is connected to a printer. If Xerox incorporated a lightweight, inexpensive printer into the base or spine of a notebook computer so that people on the go could get hard copies when and where they needed them, the company could probably win customers even if the printer wasn't as good as a stationary ink-jet printer. Only Xerox's engineers could determine whether the idea is technologically feasible. But as a strategy, this would pass the litmus tests.

If Xerox attempted this, we would expect HP to ignore this new-market disruption at the outset because the market would be much smaller than the stationary printer market. HP's printer business is huge, and the company needs large sources of new revenue to sustain its growth. To trap Hewlett-Packard in an innovator's dilemma, Xerox should develop a business model that's attractive to Xerox but unattractive to the managers of HP and other leading established printer companies. This might entail pricing ink cartridges for embedded notebook printers low enough that the executives of HP's ink jet printer business would find the market unattractive relative to investments they might make to move up-market in search of the higher profits they could find by competing against higher-cost stationary laser printers.

## 17.4.2 Conditions for Growth in Air Conditioners

The window-mounted air conditioner market is widely known to be mature, dominated by giants such as Carrier and Whirlpool. Could a company like General Electric (GE) wallop them? We would predict GE's defeat if it tried to enter this market with a quieter product that offered more features and better energy efficiency.[22] Is a low-end disruption viable? Our sense is that there are overserved customers at the low end of the existing market. They signal their overservedness by opting for the least-expensive models they can find, unwilling to pay premium prices for the alternative products that are available to them. GE might expand its already substantial manufacturing operations in China, making air conditioners for export to developed economies. This might bring modest but temporary success, because after the established companies respond by setting up their own manufacturing operations in China, GE would find itself locked in a battle with competitors whose costs are comparable and whose distribution and service infrastructure are strong, and where the targeted customers already have manifested an unwillingness to pay premium prices for better products. Employing low-cost labor constitutes a low-cost business model only until competitors avail themselves of the same option.

How about a new-market disruption, however? There are hundreds of millions of non-consumers of residential air conditioning in China, who have been blocked from that market because the power-hungry, expensive machines that historically have been available don't fit in the average family's pocketbook or apartment. If GE could design a $49.95 product that would easily slip into the window of a cramped Shanghai apartment and reduce the temperature and humidity in a ten-foot by ten-foot room with ten amps of current, things might get interesting—because once GE had a business model that could make money at that price point, taking on the rest of the up-market world would be easy. Parenthetically, while

22. Matsushita, in fact, attempted entry with a sustaining strategy of exactly this sort in the 1990s. Despite its strong Panasonic brand and its world-class capabilities in assembling electromechanical products, the company has been bloodied and has captured minimal market share.

Western executives are understandably concerned about the threat that low-cost manufacturing in China poses to them, our guess is that China's greatest competitive asset is the unfathomable amount of non-consumption in its markets, which makes them fertile ground for new-market disruptive companies of many sorts.

## 17.5 AFTERWORD

Disruption is a theory: a conceptual model of cause and effect that makes it possible to better predict the outcomes of competitive battles in different circumstances. The asymmetries of motivation chronicled in this chapter are natural economic forces that act on all businesspeople, all the time. Historically, these forces almost always have toppled the industry leaders when an attacker has harnessed them, because disruptive strategies are predicated upon competitors doing what is in their best and most urgent interest: satisfying their most important customers and investing where profits are most attractive. In a profit-seeking world, this is a pretty good bet.

Not all innovative ideas can be shaped into disruptive strategies, however, because the necessary preconditions do not exist; in such situations, the opportunity is best licensed or left to the firms that are already established in the market. On occasion, entrant companies have simply caught the leaders asleep at the switch and have succeeded with a strategy of sustaining innovation. But this is rare. Disruption does not guarantee success: It just helps with an important element in the total formula.

## 17.6 ACKNOWLEDGEMENTS

This chapter is adapted from the author's book, *The Innovator's Solution*, isbn 1578518520 [Editor's note: We highly recommend this seminal book. Published in 2003, it is still as relevant today as in 1997, and has had tremendous impact on how we think about - and practice - innovation worldwide.]

## 17.7 APPENDIX: A BRIEF DESCRIPTION OF THE DISRUPTIVE STRATEGIES OF THE FIRMS IN FIGURE 4

Table 17.2 briefly summarizes our understanding of the disruptive roots of the success of the companies that are arrayed in Figure 17.4. Because of space limitations, much important detail has been omitted. The companies are listed in alphabetical, rather than chronological, order. We do not pretend to be strong business historians, and as a consequence can only present here a partial listing of disruptive companies. Furthermore, it is often difficult to identify a specific year in which each firm's disruptive strategy was launched. Some firms existed for a considerable period, often in other lines of business, before the disruptive strategy that led to their ultimate success was implemented. In some cases it seems easier to visualize the disruption in terms of a product category, rather than by listing the name of one company. Hence, we ask our readers to regard this information as only suggestive, rather than definitive.

| Company or Product | Description |
|---|---|
| 802.11 | This is a protocol for high bandwidth wireless transfer of data. It has begun disrupting local area wireline networks. Its present limitations are that the signals can't travel long distances. |
| Amazon.com | A low-end disruption relative to traditional bookstores. |
| Apple, Compaq et.al., Personal computers | Microprocessor-based computers made by firms such as Apple, IBM and Compaq were true new-market disruptions, in that for years they were sold and used in their unique value network before they began to capture sales from higher-end professional computers. |

| | |
|---|---|
| Beef Processing | In the 1880s, Swift and Armour began huge, centralized beef slaughtering operations that transported large sides of beef by refrigerated railcar to local meat cutters. This disrupted local slaughtering operations. |
| Bell Telephone | Bell's original telephone could only carry a signal for 3 miles, and therefore was rejected by Western Union, whose business was long-distance telegraphy, because Western Union couldn't use it. Bell therefore started a new-market disruption, offering local communication – and as the technology improved, it pulled customers out of telegraphy's long distance value network into telephony. |
| Black & Decker | Prior to 1960, hand-held electric tools were heavy and rugged, designed for professionals – and very expensive. B&D introduced a line of plastic-encased tools with universal motors that would only last 25-30 hours of operation – which actually was more than adequate for most do-it-yourselfers who drill a few holes per month. In today's dollars, B&D brought the cost of these tools down from $150 to $20, enabling a whole new population to own and use their own tools. |
| Blended Plastics | These blends of inexpensive polyolefin plastics like polypropylene, sold by firms like Himont, create composite materials that in many ways share the best properties of their constituent materials. They are getting better at a stunning rate, disrupting markets that historically had been the province of engineering polycarbonate plastics made by firms like GE Plastics. |

| | |
|---|---|
| Bloomberg LP | Bloomberg began by providing basic financial data to investment analysts and brokers. It gradually has improved its data offerings and analysis, and subsequently moved into the financial news business. It has substantially disrupted Dow Jones and Reuters as a result. More recently it has created its own ECN to disrupt stock exchanges. Issuers of government securities can auction their initial offerings over the Bloomberg system, disrupting investment banks. |
| Boxed beef | The "boxed beef" model of Iowa Beef Packers completed the disruption of local butchering operations. Instead of shipping large sides of beef to local meat cutters for further cutting, IBP cut the beef into finished or nearly finished cuts, for placement directly in supermarket cases. |
| Canon photocopiers | Until the early 1980s when we needed photocopiers, we had to take our originals to the corporate photocopy center, where a technician ran the job for us. He *had* to be a technician, because the high-speed Xerox machine in there was very complicated, and needed servicing frequently. When Canon and Ricoh introduced their countertop photocopiers, they were slow, produced poor-resolution copies, and didn't enlarge or reduce or collate. But they were so inexpensive and simple to use that we could afford to put one right around the corner from our office. At the beginning we still took our high-volume jobs to the copy center. But little by little Canon improved its machines to the point that today, immediate, convenient access to high-quality, full-featured copying is almost a constitutional right in most workplaces. |

| | |
|---|---|
| Catalog retailing | Sears, Roebuck and Montgomery Ward took root as catalog retailers – enabling people in rural America to buy things that historically had not been accessible. Their business model, entailing annual inventory turns of 4x and gross margins of 30%, was disruptive relative to the model of full-service department stores, which relied upon 40% gross margins because they turned inventories only 3x annually. Sears and Wards later moved up-market, building retail stores. |
| Charles Schwab | Started in 1975 as one of the first discount brokers. In the late 1990s Schwab created a separate organization to build an on-line trading business. It was so successful that the company shut down its original organization of telephone brokers. |
| Circuit City, Best Buy | Disrupted the consumer electronics departments of full-service and discount department stores, which has sent them up-market into higher-margin clothing. |

| | |
|---|---|
| Cisco | Cisco's router uses packet-switching technology to direct the flow of information over the telecommunications system, compared to the circuit-switching technology of the established industry leaders such as Lucent, Siemens and Nortel. The technology divides information into virtual "envelopes" called packets, and sends them out over the Internet. Each packet might take a different route to the addressed destination; and when they arrive, the packets are put in the right order and "opened" for the recipient to see. Because this process entailed a few seconds' latency delay, packet switching could not be used for voice telecommunications. But it was good enough to enable a new market to emerge – data networks. The technology has improved to the point that today, the latency delay of a packet-switched voice call is almost imperceptibly slower than that of a circuit-switched call – enabling VOIP, or voice-over-Internet-protocol telephony. |
| Community colleges | In some states, up to 80% of the graduates of reputable four-year state universities took some or all of their required general education courses at much less expensive community colleges, and then transferred those credits to the university – which (unconsciously) is becoming a provider of upper-division courses. Some community colleges have begun offering four-year degrees. Their enrollment is booming, often with non-traditional students who otherwise would not have taken these courses. |

| | |
|---|---|
| Concord School of Law | Founded by Kaplan, a unit of the Washington Post Company, this on-line law school has attracted a host of (primarily) non-traditional students. The school's accreditation allows its graduates to take the California Bar exam, and its graduates' success rate is comparable to those of many other law schools. Many of its students don't enroll to become lawyers, however. They want to understand law to help them succeed in other careers. |
| Credit scoring | A formulaic method of determining creditworthiness, substituting for the subjective judgments of bank loan officers. Developed by a Minneapolis firm, Fair Isaac. Used initially to extend Sears and Penny's in-store credit cards. As the technology improved, it was used for general credit cards, and then auto, mortgage and now small business loans. |
| Dell Computer | Dell's direct-to customer retailing model and its fast-throughput, high asset-turns manufacturing model allowed it to come underneath Compaq, IBM and Hewlett Packard as a low-end disruptor in personal computers. Clayton Christensen, the quintessential low-end consumer, wrote his doctoral thesis on a Dell notebook computer purchased in 1989, because it was the cheapest portable computer on the market. Because of Dell's reputation for marginal quality, students needed special permission from Harvard to use doctoral stipend money to buy a Dell rather than a computer with a more reputable brand. Today Dell supplies most of the Harvard Business School's computers. |

| | |
|---|---|
| Department Stores | Department stores like Z.C.M.I. in Salt Lake City, Marshall Field in Chicago, and R.H. Macy in New York, disrupted small shopkeepers. The department stores made money by accelerating inventory turns to 3x per year, which enabled them to earn attractive profit with 40% gross margins. Because their salespeople were much less knowledgeable about products, at the outset department stores had to start at the simplest end of the merchandise mix, with products that were so familiar in use that they "sold themselves." |
| Digital animation | The fixed cost and skill required to make a full-length animated movie historically was so high that almost nobody could do it except Disney. Digital animation technology now enables far more companies (Such as Pixar) to compete against Disney. |
| Discount department stores | Department stores like Korvette's in New York, K-Mart in Detroit, and later Wal-Mart and Target disrupted full-service department stores. The discount stores made money by accelerating inventory turns to 5x per year, which enabled them to earn attractive profit with 23% gross margins. Because their salespeople were much less knowledgeable about products, at the outset the discount department stores had to start at the simplest end of the merchandise mix, with branded hard goods that were so familiar in use that they "sold themselves." They subsequently have moved up-market into soft goods such as clothing. |

| | |
|---|---|
| E-Bay | Most of the Internet start-ups of the late 1990s attempted to use the Internet as a sustaining innovation relative to the business models of established companies. E-Bay was a notable exception, as it pursued a new-market disruptive strategy – enabling owners of collectibles that could never turn the head of auction house executives, now to be able to sell off things that they no longer needed. |
| ECNs | Electronic clearing networks (ECNs) allow buyers and sellers of equities to exchange them over a computer, at a fraction of the cost of doing it on a formal stock exchange. Island, one of the leading ECNs, can handle on one workstation volume amounting to 20% of the NASDAQ's volume. |
| E-mail | E-mail is disrupting postal services around the globe. The volume of personal communication that is done by letter is dropping precipitously, leaving postal services with magazines, bills and junk mail. |
| Embraer & Canadair regional jets | The regional passenger jet business is booming, as their capacity over the past 15 years has stretched from 30 to 50, 70 and now 106. As Boeing and Airbus struggle to make bigger, faster jets for transcontinental and transoceanic travel, their growth has stagnated; the industry has consolidated (Lockheed and McDonnell Douglas have been folded in); and the growth is at the bottom of the market. |

| | |
|---|---|
| Endoscopic Surgery | Minimally invasive surgery was actively disregarded by leading surgeons because the technique could only address the simplest procedures. But it has improved to the point that even certain relatively complicated heart procedures are done through a small port. The disruptive impact has primarily been on equipment makers and hospitals. |
| Fidelity management | Created "self-service" personal financial management through its easy-to-buy families of mutual funds, 401k accounts, insurance products, etc. Fidelity was founded a few years after WWII; but began its disruptive movements in the 1970s, as best we can tell. |
| Flat panel displays (Sharp et.al.) | We normally think of disruptive technologies as being inexpensive, and many people are puzzled at how we could call flat panel displays disruptive. Haven't they come from the high end? Actually, no. Flat panel LCD displays took root in digital watches; and then moved to calculators, notebook computers and small portable televisions. These were applications that historically had no electronic displays at all, and LCD displays were *much* cheaper than alternative means of bringing imaging to those applications. Flat screens have now begun invading the mainstream market of computer monitors and in-home television screens, disrupting the cathode ray tube. They are able to sustain substantial premium prices because of their 2-D character. |

| | |
|---|---|
| Ford | Henry Ford's Model T was so inexpensive that he enabled a much larger population of people who historically could not afford cars, now to own one. |
| Galanz | China's Galanz captured nearly 40% of the world microwave oven market in the 1990s. While the company could have followed a strategy of low-end disruption – using low-cost Chinese labor to make appliances for export, it instead chose to be a new-market disruptor, making ovens that were small enough and consumed little-enough power to be used in cramped Chinese apartments; and were cheap enough for non-microwave oven owners to afford. Once they had built a business model that could make market-enabling price points for the domestic Chinese market, then taking on the rest of the world was as easy as egg-drop soup. |
| GE Capital | Has disrupted major portions of the commercial banks' historical markets, primarily through low-end disruptive strategies. |
| Google | Google and its competing Internet search engines are disrupting directories of many sorts, including the Yellow Pages. |
| Honda motorcycles | Honda's Supercub, introduced in the late 1950s, disrupted makers of big, thunderous motorcycles such as Harley Davidson, Triumph, BMW and many others. It took root as an off-road recreational motorized bicycle, and then improved. Honda was joined by Yamaha, Kawasaki and Suzuki. |

| | |
|---|---|
| Ink jet printers | These were a disruption to the laser jet printer, and a sustaining technology relative to the dot-matrix printer. We put ink jet printers toward the "new market" end of the disruption spectrum, because their compact size, light weight and low initial cost enabled a whole population of computer owners – primarily students – each to own and use a printer. While they were slow and produced fuzzy images at the outset, ink jet printers are now the mainstream printer of choice, having pushed laser jets to the high end. Hewlett Packard stayed atop this industry by setting up an autonomous ink jet business unit to compete against its laser jet printer business. |
| Intel micro-processor | Intel's earliest microprocessor in 1971 could only constitute the brain of a four-function calculator. Makers of computers whose logic circuitry is microprocessor-based have disrupted firms that made mainframe and minicomputers, whose logic circuitry was printed wiring board-based. |
| Intuit's *QuickBooks* accounting software | Whereas the established industry leaders in accounting software enabled small business managers to run all sorts of sophisticated reports for analytical purposes, *QuickBooks*, which was a derivative of Intuit's personal finance software product *Quicken*, basically helped them keep track of their cash. It created a huge new market amongst very small business owners (most less than five employees) who historically did not keep their books on computer. Within two years of launch Intuit had seized 85% of the small business accounting software market – mainly by creating new growth. The stealing of the established companies' customers came later, as *QuickBooks*' functionality improved. |

| | |
|---|---|
| Intuit's *Turbotax* | PC-based accounting software is disrupting personal tax preparation services such as H&R Block. |
| Japanese Steel Makers | Firms like Nippon Steel, Nippon Kokkan and Kobe and Kawasaki Steel began their growth by exporting very low quality steel to western markets starting in the late 1950s. As their customers (including disruptive Japanese auto makers like Toyota) grew, the Japanese steel industry had to increase capacity dramatically, enabling it to incorporate the latest steelmaking technology like continuous casting and basic oxygen furnaces in the new mills. This accelerated their up-market trajectory dramatically. |
| Jet Blue | Whereas Southwest Airlines initially followed a strategy of new-market disruption, Jet Blue's approach is low-end disruption. Its long-range viability depends upon the major airlines' motivation to run away from the attack, as integrated steel mills and full-service department stores did. |
| Kodak | Until the late 1800s, photography was extremely complicated. Only professionals could own and operate the expensive equipment. George Eastman's simple "point and shoot" "Brownie" camera allowed consumers to take their own pictures. They could then mail the encased roll of film to Kodak, which would develop and return the photos by mail. |

| | |
|---|---|
| Kodak Funsaver | Kodak's *Funsaver*-brand single-use camera was born out after painful labor within Kodak, because its profit model – gross margins – were lower than Kodak could earn by selling roll film; and the quality of the images was not as good as those taken in high-quality 35mm cameras. But Kodak commercialized it through a different division, and it sold almost exclusively to people who would not have bought film anyway – because they didn't have a camera. While it has potential to move up-market taking share against traditional cameras with a new brand, *Maxx*, we sense that Kodak has stopped driving it in this direction. |
| Korean auto manufacturers: Hyundai & Kia | Korean automakers, including Hyundai and Kia, gained more points of worldwide market share in the 1990s than any other country's automakers. And yet few of the established firms are concerned, because their gains have come in what is, to them, the lowest-profit portion of the market. |
| MBNA | We noted above that credit scoring is a formulaic method of determining the creditworthiness of a loan applicant. It was originally implemented in commercial banks as a sustaining technology – to reduce their costs of credit evaluation. In the 1990s, however, it was deployed in high-volume, low-cost "monoline" business models by firms such as MBNA, Capital One and First USA, which have substantially disrupted commercial banks' credit card business. At the time of this writing, in fact, Citibank is the only major commercial bank with a substantial and profitable credit card business. |

| | |
|---|---|
| McDonald's | The fast food industry has been a hybrid disruptor, making it so inexpensive and convenient to eat out that they created a massive wave of growth in the "eating out" industry. Their earliest victims were "mom-and-pop" diners. In the last decade the advent of food courts has taken fast food up-market. Expensive, romantic high-end restaurants still thrive at the high end, of course. |
| MCI, Sprint | These firms were low-end disruptors relative to AT&T's long distance telephone business. They enjoyed a unique opportunity to do this, because AT&T's long distance rates were set by regulation at artificially high levels, in order to subsidize local residential telephone service. |
| Merrill Lynch | Charles Merrill's mantra in 1912 was to "Bring Wall Street to Main Street." By employing salaried rather than commissioned brokers, he made it inexpensive enough to trade stocks that middle-income Americans could become equity investors. Merrill Lynch moved up-market over the next 90 years towards higher net worth investors. Most of the brokerage firms that held seats on the New York Stock exchange in the 1950s and 60s have been merged out of existence, because Merrill Lynch disrupted them. |
| Microsoft | Its operating system was inadequate versus those of mainframe and minicomputer makers; versus Unix; and versus Apple's system. But its migration from DOS to Windows to Windows NT is taking the firm up-market, to the point that the Unix world is seriously threatened. Microsoft, in turn, faces a threat from Linux. |

| | |
|---|---|
| Mini-computers | Companies like Digital Equipment, Prime, Wang, Data General and Nixdorf were new-market disruptors relative to mainframe computer makers. Their relative simplicity and low price enabled departments (particularly engineering) in organizations to have their own computers, instead of having to rely on inconvenient, centralized mainframe computers that typically were optimized for generating financial reports. |
| On-line stock brokers | On-line trading of equities is a sustaining technology relative to the business models of discount brokers such as Ameritrade, and is disruptive relative to full-service brokers such as Merrill Lynch. For Schwab, which started as a bare-bones discount broker but had moved up towards the mainstream market by the mid-1990s, Internet-based trading was disruptive enough that the company had to set up a separate division. |
| On-line travel agencies | Enabled by electronic ticketing, on-line travel agencies such as Expedia and Travelocity have so badly disrupted full-service, bricks-and-mortar agencies such as American Express that many airlines have dramatically cut the substantial commissions that historically they had paid to travel agencies. |

| | |
|---|---|
| Oracle | Oracle's relational database software was disruptive relative to that of the prior leaders, Cullinet and IBM, whose hierarchical or transactional database software ran on mainframe computers and was used to generate standard financial reports. Relational databases ran on minicomputers (and then microprocessor-based computers). Users without deep programming expertise could readily create their own custom reports and analyses using Oracle's modular, relational architecture. |
| Palm Pilot, RIM BlackBerry | Hand-held devices are new-market disruptions relative to notebook computers. |
| Plastics | Plastics as a category have disrupted steel and wood, in that the "quality" of plastic parts often was inferior to those of wood and steel, along the metrics by which performance was measured in traditional applications. But their low cost and ease of shaping created many new applications, and plastics have pulled many applications out of the original metal and wood value networks into the plastic network. The disruption is particularly obvious if you look at where plastics were used in automobiles 30 years ago, versus today. |
| Portable diabetes blood glucose meters | Disrupted makers of large blood glucose testing machines in hospital laboratories, enabling patients with diabetes to monitor their own glucose levels. |
| Salesforce.com | This company, with its inexpensive, simple Internet-based system, is disrupting the leading providers of customer relationship management software like Siebel Systems. |

| | |
|---|---|
| Seiko watches | Remember when Seiko watches were those cheap, throw-away black plastic watches? They, Citizen and Texas Instruments (which subsequently exited) disrupted the American and European watch industries. |
| Sonosite | This firm makes a hand-held ultrasound device that enables healthcare professionals who historically needed the assistance of highly trained technicians with expensive equipment, now to look inside the bodies of patients in their care, and thereby to provide more accurate and timely diagnoses. The company floundered for a time attempting to implement its product as a sustaining innovation. But as of the time this book was being written, it seemed to have caught its disruptive stride in an impressive way. |
| Sony | Sony pioneered the use of transistors in consumer electronics. Its portable radios and portable televisions disrupted firms like RCA that made large TVs and radios using vacuum tube technology. During the 1960s and 1970s, Sony launched a series of new-market disruptions, with products like video tape players, hand-held consumer video recorders, cassette tape players, the Walkman, and the 3.5-inch floppy disk drive. |
| Southwest Airlines | It was a hybrid disruptor because its original strategy was to compete against driving and busses, and to fly in and out of non-mainstream airports. In addition, because its prices were so low it also took business from established airlines. Just as Wal-Mart enjoys profit protection from being in small towns whose market can only support one discount store, many of Southwest's routes offer the same protection. |

| | |
|---|---|
| SQL database software | Microsoft's SQL database software product is disrupting Oracle, which has moved up-market into expensive, integrated enterprise systems. |
| Staples | With its direct competitors Office Max and Office Depot, Staples disrupted small stationery stores as well as business-to-business office supplies distributors. |
| Steel minimills | Have been disrupting integrated mills around the world since the mid-1960s, as recounted in the text. |
| Sun Microsystems | Sun, Apollo (HP) and Silicon Graphics, which built their systems around RISC microprocessors, took root in essentially the same value network as minicomputers, and disrupted them. These firms, in turn, are now being disrupted by CISC microprocessor-based computer makers such as Compaq and Dell. |
| Toyota | Entered the US market with cheap sub-compact cars like the Corona. These were so inexpensive that people who historically couldn't afford a new car now could buy one; or families could acquire a second car. Toyota now makes Lexuses, you may have noticed. Nissan has migrated from its Datsun to Infiniti; and Honda has progressed from its miniature CVCC to Accura. |
| Toys-R-Us | Disrupted the toy departments of full-service and discount department stores, which has sent them up-market into higher-margin clothing. |

| | |
|---|---|
| Ultrasound | Ultrasound technology is disruptive, relative to X-Ray imaging. Hewlett Packard, Accuson, and ATL created a multi-billion-dollar industry by imaging soft tissues, which traditional X-ray technology could not capture. The leading X-Ray equipment makers, including General Electric, Siemens and Philips, became leaders in the two major radical sustaining technology revolutions in imaging: CT scanning and magnetic resonance imaging (MRI). Because ultrasound was a new market disruption, none of the X-ray companies participated in ultrasound until very recently, when they acquired major ultrasound equipment companies. |
| University of Phoenix | A unit of Apollo, the University of Phoenix is disrupting four-year colleges and certain professional graduate programs. It began by providing employee training courses for businesses, often *de facto*, but sometimes by formal contract. Its programs have expanded into a variety of open-enrollment, degree-granting programs. Today it is one of the largest educational institutions in the United States, and is one of the leading providers of on-line education. |
| Unmanned aircraft | These machines took root initially as drone targets to uncover hidden anti-aircraft emplacements. They then moved up-market into surveillance roles, and in the 2001-02 war in Afghanistan, moved for the first time into limited weapons-carrying roles. |
| Vanguard | Index mutual funds have been a low-end disruption relative to managed mutual funds. At the time of this writing, Vanguard's assets had grown to rival closely those of the former undisputed mutual fund leader, Fidelity management. |

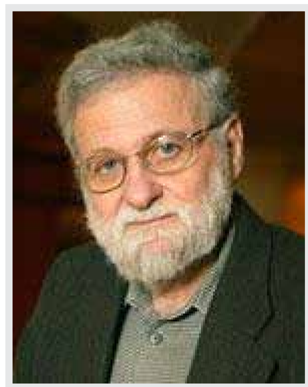
| | |
|---|---|
| Veritas & Network Appliance | Network-attached storage and IP storage area networks are disruptive approaches to enterprise data storage, relative to the centralized storage systems supplied by companies like EMC. Some of these distributed networked storage are so simple to augment that an office assistant can simply "snap" an additional storage server onto a network. |
| Wireless Telephony | Cellular and digital wireless phones have been on a disruptive path against wireline phones for 25 years. Initially they were large, power-hungry car phones with spotty efficacy; but gradually have improved to the point where, by some estimates, nearly one-fifth of mobile telephone users have chosen to "cut the cord" and do without wireline telephone service. The viability of the wireline long distance business is now in jeopardy. |
| Xerox | Photocopying has been a new-market disruption relative to offset printing, enabling non-printers to make smaller volumes of copies in the convenience of their workplace. Xerox's initial machines were so expensive and complicated that they were housed in corporate photocopy centers manned by technicians. |

Table 17.2: A Brief Description of the Disruptive Roots of the Companies and Industries Listed in Figure 4.

## 17.8 COMMENTARY BY DONALD A. NORMAN

How to cite this commentary in your report

### Donald A. Norman



© *Donald A. Norman*

Don Norman is the author of numerous books including "Emotional Design," and more recently, "Living with Complexity." He is co-founder of the Nielsen Norman group, a professor at KAIST (in Korea), and IDEO fellow, and a design theorist, studying the fundamentals of modern design.Donald A. Norman has a Bachelor of Science in Electrical Engineering from MIT and a Doctor of Philosophy in Ps...

Donald A. Norman
Donald A. Norman is a member of The Interaction Design Foundation

Anyone who cares about innovation must read Clay Christensen. Why? Let me start with some history.

I first encountered Christensen's works when I was at HP, in 1997. His first book was still in manuscript form and was widely circulated among a small group of enthusiastic managers. I got a copy and also fell in love. Most importantly, it precisely described the situation we were in at HP: we had several disruptive products in the pipeline, but the executives at HP were incredibly risk aversive, so they shunned them, or in some cases, required them to be so watered down and deprived of resources for proper development, that they became self-fulfilling prophecies for the executives: they were doomed to fail.

We even brought Christensen to HP. I remember well a talk he gave which covered our situation precisely. The reception by the audience was wonderful: the reception by HP executives was dismal. Afterwards a few of us gathered around him and told him that we were case studies of the kinds of failure he was describing.

When I wrote my book "The Invisible Computer," I used Christensen's work as a starting point for my discussion of the book's subtitle: "Why good products can fail." I modified his basic graph (see Figure 1 of Christensen's article) to reframe the point in design terms. The result, shown here, as my Figure 1, should be self-explanatory.

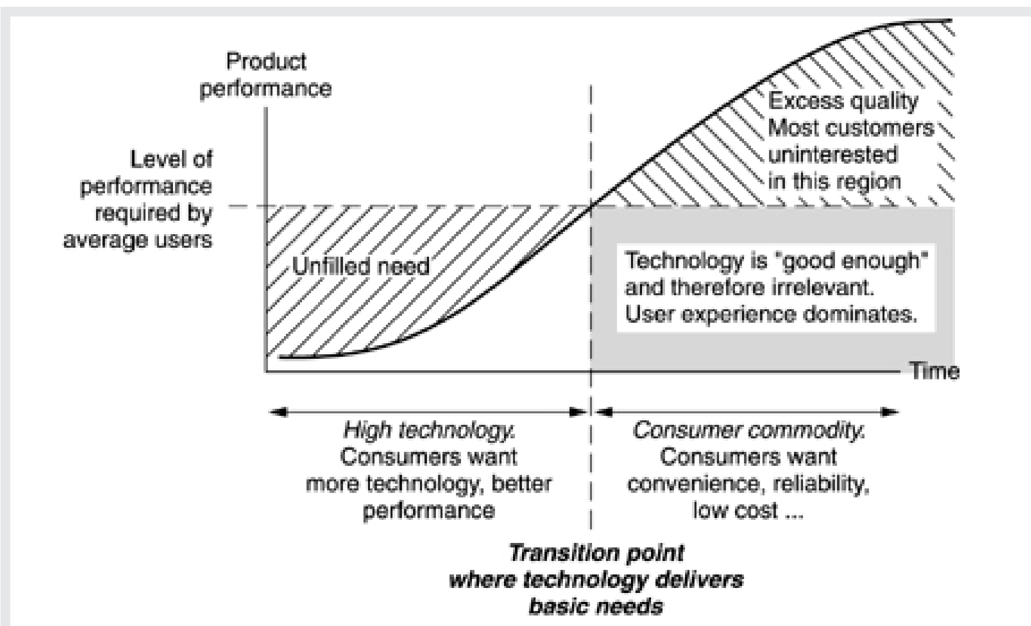



Figure 17.1: The needs-satisfaction curve of a technology. New technologies start out at the bottom left of the curve: delivering less than the customers require. As a result, customers demand better technology and more features, regardless of the cost or inconvenience. A transition occurs when the technology can now satisfy the basic needs. Figure 2.2 of Norman (1998), modified from Christensen (1997).

My book was about the way new products get adopted by the market. The standard view, that market acceptance starts with early adopters and then, slowly, brings in late adopters, was first formulated in 1962 by the Stanford professor Everett Rogers (1995) and then publicized in Geoffrey Moore's book "Crossing the Chasm" (Moore, 1995: alas, in this book Moore failed to give credit to Rogers, an omission he corrected in his next book). These two groups of adopters are very different. Indeed, Moore argued that they were separated by a chasm that could only be bridged by a better product and different marketing. I often describe the difference by stating that for early adopters, the technological promise suffices. For late adopters, human-centered design is essential, for these people don't want promises, they want easy to understand, effective, enjoyable products. I realized that this view of the product acceptance cycle could be combined with Christensen's insights of the relationship of technological capabilities to customer needs: Figure 2.

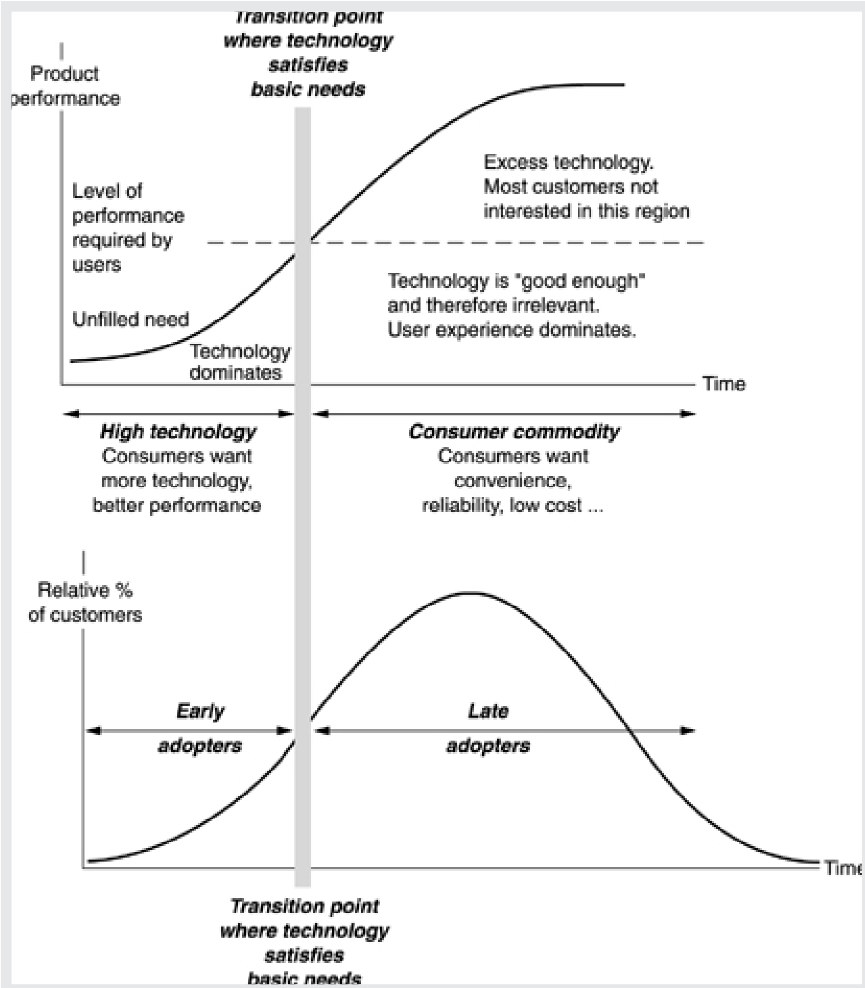



Figure 17.2: The change from technology-driven products to customer-driven, human-centered ones. As long as the technology's performance, reliability and cost falls below customer needs, the marketplace is dominated by early adopters: those who need the technology and who will pay a high price to get it. But the vast majority of customers are late adopters. They hold off until the technology has proven itself, and then they insist upon convenience, good user experience, and value. Figure 2.4 of Norman (1998).



Since the publication of his first book, Christensen has continued to expand and elaborate upon his ideas of product disruption. In the original formulation of his model, including my adaption, the line that depicts "required performance" is horizontal. But as Christensen points out in his article for this encyclopedia, the level should differ for different people and different applications, and so in his later work he has reminded readers that the line is simply the average of the user base. More importantly, he points out that the line should not be horizontal. After all, with the passage of time and the acceptance of the technology, people's needs change, becoming more sophisticated and requiring more performance. At the least, therefore, the line must slope upwards, so that with the passage of time, more is required than in the early stages. Companies try to accommodate this changing need through the addition of extra features, more powerful processors, and more powerful everything, from processors, to motors, to whatever variables are of interest. These additions and minor modifications of Christensen's model do not change the basic thrust.

Christensen has also greatly strengthened this model by adding a new dimension: market. In the original model depicted by Figure 1, product performance is plotted against time. In this two-dimensional model, the only enhancements relevant are price and performance. Here, disruptive innovation always occurs at the low end of performance: new entrants to the market provide less performance, but a very low cost, thus satisfying the needs of people who do not require the ever-expanding capabilities of the main products.

The modified model recognizes that disruption can occur by changing the market population, serving people who would never would have even considered the product. Hence the third dimension: market. Want an example? Consider the first home computers, for example the Apple II. These were puny devices, no match in quality for the computers then in use in universities and companies. They offered no challenge to the makers of these machines, which meant their makers could safely ignore them. But individuals and students could never have purchased any of the large, complex, and expensive computers that were then being sold to industry. As a result, the Apple II opened up an entire new market of customers: individuals bought them and fell in love with their capabilities, even if they were very limited and of low quality. Eventually, these weak, puny home computers took over the home market while slowly expanding their capabilities. Eventually, they became today's powerful machines that now dominate in home, business, education, and entertainment.

I can attest to this issue, for when I first experienced the Apple II computer, I wondered why anyone would ever buy one. To me, it was a huge step backwards from the laboratory machine I was using all day at the University. A year or two later I surprised myself by buying one for my family, and years after that, I left the university and went to work for Apple. (A truly disruptive technology.) These early machines opened up an entire new marketplace for people who would never have thought of purchasing an IBM office computer or a DEC machine (Digital Equipment Corporation, later known as Digital), the two companies that then dominated the world of computers. Over time, of course, the home computer took over the world. DEC no longer exists. SUN no longer exists. Silicon Graphics no longer exists. And IBM has gotten out of the personal computer business.

So, the modified model has two ways by which disruptive influence can take place: low-end disruption or market disruption.

Note that it would be a mistake to equate these arguments with different kinds of innovative forces: these are not the same as incremental or radical disruption, as Christensen points out.

### 17.8.1 The theory is easy to understand: the practice is extremely difficult

It is really easy to understand the message that Christensen presents. It is really difficult to execute upon that message. Why?

Consider the story of Eastman Kodak. Long the world's leader in photography, it has faltered. What happened it was disrupted by the rise of digital photography. Why? Didn't they see it coming? No, they missed the early signs because their customers all wanted high quality films that delivered high quality images: digital could not even come close to competing.

Kodak made cameras, but its real income was derived from film and chemicals. Serious amateurs and professionals all preferred film. Radiologists explained at great length why only film X-Rays could deliver the depth of contrast and resolution they demanded. Professional photographers said the same thing. The customers drove the old industry.

Kodak knew about digital technology. Their researchers were working on it. They had produced one of the first digital cameras (a joint project with Apple). And they even hired as CEO someone from HP who was a champion of digital photography. But it was too late. Moreover the strength of the company was in analog technology, in imaging, in film, and in chemistry. Imaging was still relevant: the rest was much less relevant.

And the customers were correct. The early digital cameras were inferior. I am fond of giving talks about my failures. One of the first projects I watched when I joined Apple was a failure. It was an exciting, well-done product, but it failed in the marketplace. What was it? A digital camera (the joint project with Kodak). Why did it fail? See Figure 1. It was far to the left. Basically, products that are too early will fail. The customers aren't ready and the technology is not good enough. Our camera did not have a display screen (they were too expensive then), it took low-resolution photos, and it could only hold a few of them. It was difficult to transfer the photos to a PC and then difficult to work with them. Inexpensive ink-jet printers did not exist, so there was no way to print them in color.

Apple withdrew the camera from the market. Too bad: had Apple stuck to it, it would have been the leader in digital photography instead of the poor humble company that it is now. (Self protection: that last sentence is meant as a joke.)

When you are a historian looking back, it is easy to see the disruptive technology and wonder why the existing companies do not jump on them. When you are inside the companies, it is very difficult to see. Inside the company it looks like yet another one of those hair-brained research projects by those impractical researchers that is completely impractical.

I can give other examples from my own experience. I, and my co-workers at both Apple and HP, would sometimes identify new ideas and emerging products as disruptive forces that would change the industry, only to discover that they didn't, either because they were simply the wrong idea or, in the case of some things (such as the digital camera), it was simply premature. Another example would be or Apple's Newton, which, if you never heard of it, simply proves my point. Newton contained many of the innovations that today are commonplace in our smart phones, but it died after a very public, very humiliating life experience.

Similarly, my co-workers and I dismissed radical new products as irrelevant, such as my early dismissal of the personal computer or Apple's dismissal of the first browser (a long story, to be told some other time). The view of a company from the perspective of an academic scholar, looking back in time over the historical record is very different from the view of the company itself. For that matter, the view is from different from the executive suite than from that seen by the middle managers, or individual contributors. The view from the company's research laboratories is different once again. And finally, the view from the member of a small, just-funded startup company is yet again very different. The large company is biased not to see disruptions: the startup is biased not to see hurdles that might prevent them from causing an immense disruption.

Christensen's analyses are important and influential. I, myself, find them very compelling. But they are very difficult to apply. After reading Christensen's chapter I thought about each of the nine companies I am currently advising, attempting to see how the information in the chapter would be relevant. Some are small startups, but others are larger, including at least one of the largest companies in the world. In every single case I concluded that the company was different and didn't fit the mold. This was true even for the several companies that believe they are about to cause a major upheaval in product space. Yes, they might be disruptive, I concluded, but not in the sense described in these books for they were entering new territories where comparative products do not exist. Is my analysis correct? Only time will tell.

Although I use business cases in teaching students, these cases are written long after the event. They are written with the benefit of hindsight, for the writers know what actually happened. As a result, the case always depicts the events in a manner that is cleaner and more logical than the experience of those living the case. A case history invariably ends with a question: which option should the company select? Well, I have lived through the examples given in some of those cases. It didn't seem that clean and logical to us. At the time, we didn't recognize that we were at a decision point: we didn't recognize those nice, clearly stated options.

Analysis is easier than syntheses. Hindsight is easier than foresight. The view from inside the battle is different from that of the historian.

### 17.8.2 Comment on the Chapter

Note that the chapter presented here is an excerpt/adaptation from the Christensen's book "The Innovator's Solution," which was published in 2003. As a result, many of the examples are dated. This is especially the case for Table 17.2 of the Appendix, where the examples of companies are often quite radically out of date. I will leave the analysis of the today's relevance of the companies to the points of the article as an exercise for the reader. Just realize that the table was constructed over a decade ago, which in the case of some businesses is a lifetime. Indeed, given the rate at which technology businesses are absorbed by other companies or simply fail, it can be more than a lifetime.

### 17.8.3 References


- Christensen, C. M. (1997). The innovator's dilemma: When new technologies cause great firms to fail. Boston: Harvard Business School Press.
- Moore, G. A. (1991). Crossing the chasm: Marketing and selling high-tech goods to mainstream customers. New York: HarperBusiness.

- Norman, D. A. (1998). The invisible computer: why good products can fail, the personal computer is so complex, and information appliances are the solution. Cambridge, MA: MIT Press.
- Rogers, E. M. (1962). *Diffusion of innovations*. New York: Free Press. (Five editions were published, the last being in 2003.)

### About Don Norman

Don Norman wears many hats, including co-founder of the Nielsen Norman group, IDEO fellow, Visiting Professor at KAIST (South Korea), consultant, advisor, board member, and author. His latest book is *Living with Complexity*. He lives at jnd.org.

## 17.9 COMMENTARY BY MARC STEEN

How to cite this commentary in your report

### Marc Steen

© *Marc Steen*

Marc Steen works as a senior scientist at research and innovation organization TNO in the Netherlands. He earned MSc, MTD and PhD degrees in Industrial Design Engineering at Delft University of Technology and has worked at Philips and

KPN before joining TNO. His expertise is in human-centred design, co-design, open innovation and innovation management. His research interests include p...

Marc Steen
Marc Steen is a member of The Interaction Design Foundation

### 17.9.1 A social perspective: On empowerment, flourishing, cooperation and creativity

Christensen discusses the concept of disruptive innovation in detail and with eloquence. He convincingly argues that most firms tend to focus on sustaining innovations—incrementally improving their products and services, aiming to serve the attractive higher end of their current customer base—and that they thus unintentionally, create opportunities for new entrants or new ventures of other firms. These new entrants or new ventures can introduce new products or services at the lower end of the market—offering a 'good enough' product for lower costs and for lower prices, serving people that are currently over-served (*low-end disruptions*), or offering 'good enough' products or services for new customers or new situations for consumption or usage (*new-market disruptions*). These new entrants or new ventures start at the lower end of a market and progressively move up through this market, and gradually conquer the established firms' businesses. Not by attacking them head-on, but by first taking a piece of the cake that nobody is really interested in, and then a next piece, and a next piece...

In this chapter (above), Christensen adopts a business perspective and focuses on economics and market dynamics. But for those who feel less comfortable with such economic and commercial vocabulary, you may rest assured that disruptive innovation is also about empowerment and promoting development, freedom and well-being, and about promoting processes of design thinking, cooperation and creativity.

Below, I will adopt a social perspective on disruptive innovation. I am not the first to do this. Christensen and his colleagues have discussed the role of disruptive innovation in bringing about positive social change (Christensen et al., 2006), focusing on education (Christensen et al., 2008) and health care (Christensen et al., 2009). They discuss examples of entrepreneurship in education in health care in which ‘good enough’ services—based on relatively cheap processes, for example, by using ICT—are provided to people who can not afford or use current services. For example, by offering affordable and reliable basic health advice in convenience stores or by offering high quality and cost-effective online courses for distance education.

### 17.9.2 Empowering people at the ‘base of the pyramid’ to flourish

This reminds me of *Design for the real world* (Papanek, 1991), the book that first made me think about the roles that designers can play in bringing about positive social change, by focusing on developing products and services that meet real needs of real people, rather than producing more stuff for the affluent.

And it reminds me of serving the ‘bottom of the pyramid’ (Prahalad, 2004), which refers to the provisioning of products or services to large groups of relatively poor people—typically in developing countries—in order to both support these people to flourish, and to enable companies to make money by offering these products and services. This is done, preferably, by promoting local and social entrepreneurship, so that ‘poor’ people can become producers and partners (Immelt et al. 2009), rather than be treated as receivers or consumers. Promoting well-being, sustainable economic development and commercial success go hand in hand in this approach. A relatively large portion of our attention typically goes to serving the top of the pyramid, rather than serving the base. Individual people at the base may not have much to spend but their large number makes them an interesting target group. Developing products or services for the base of the pyramid is often a good example of disruptive innovation, since these are typically produced for lower costs and sold for lower prices than existing products or services.

This notion of serving the base of the pyramid (BoP) can be further developed using the *capability approach* (Sen, 1999; Nussbaum, 2011; see also Dong, 2008; Oosterlaken, 2009), which presents a framework to evaluate to what extent a specific BoP project actually contributes to people's development and social change.

The capability approach was developed in order to design and evaluate well-being and development programmes in developing countries, and focuses on the empowerment of people—where empowerment is understood as an increase of certain important capabilities. The capability approach offers an alterative to approaches that focus on providing specific commodities (such as water wells or computers) in that it acknowledges that these commodities can only be used if a range of personal factors (such as personal skills), social factors (such as social norms) and environmental factors (such as infrastructure) are in place. It also offers an alternative to approaches that focus on promoting specific behaviours (such as using machines in specific ways) in that it acknowledges that people should have freedom to decide for themselves how they want to live their version of 'the good life', how they want to flourish. Freedom and development are intrinsically and intimately intertwined in the capability approach (Sen, 1999).

Those that are interested in bringing about positive social change can get inspiration from BoP and the capability approach literature, which suggest various ways to empower people to improve their capabilities—and to increase their well-being.

### 17.9.3 Design thinking, cooperation and creativity in public services

Disruptive innovation also reminds me of the application of design thinking in public services innovation (Brown & Wyatt, 2010; Thomas, 2008). People, for example, in the UK, found ways to empower citizens to co-create or co-produce public services (Cottam & Leadbeater, 2004; Boyle & Harris, 2009). In health care, one would, for example, promote relatively cheap, bottom-up self-help, informal care and prevention activities by citizens, rather than depend too much on relatively expensive, top-down care by professionals. Design thinking is applied to rethink and redesign public services— focusing on participation, cooperation and creativity—helping to develop services that are often cheaper to produce and offer the same or even higher social value for those involved.

Such efforts draw from diverse design disciplines and apply diverse design methods, perspectives and approaches to the development and implementation of public services. For example participatory design (Schuler & Namioka, 1993; Muller, 2002) or co-creation (Sanders & Stappers, 2008) methods, such as workshops in which citizens, civil servants and others jointly discuss problems and jointly develop solutions. Or service design (Parker & Heapy, 2006) perspectives, which focus on the needs and experiences of those people whom are most involved in the service, for example, exploring patients' needs and nurses' expertise and developing services that create a mach between these two. Or transformation design (Burns et al., 2006) approaches, in which people with different backgrounds jointly define the problem and jointly explore possible solutions at a systems-level, rather than at the micro-level of an individual organization—thus enabling radical and systemic innovations, rather than developing local or micro solutions, which can be sub-optimal. Moreover, transformation design helps the organizations involved to improve their capabilities for innovation, cooperation and creativity, so that these become integral parts of their ways of working.

In all these methods, perspectives and approaches the active and creative participation of citizens, as clients or users of public services, and of civil servants, as providers of these public services, and cooperation and joint creativity between people from different organizations, with different backgrounds are critical. Ideally, these methods, perspectives and approaches are concrete manifestations of a process of *design thinking*, a process which 'involves finding as well as solving problems' (Lawson, 2006: p. 125) so that the 'problem and solution co-evolve' (Cross, 2006: p. 80). Ideally, diverse people participate in a process in which they jointly explore and articulate the problem, and explore and develop possible solutions, in an iterative process.

Those that are interested in improving or redesigning public services can get inspiration from design thinking literature—on co-design, service design and transformation design—in order to more effectively organize cooperation and creativity.

Like Christensen et al. (2006; 2008; 2009), I believe that disruptive innovation can have a key role in promoting positive social change, by empowering people to flourish and by promoting cooperation and creativity.

## 17.9.4 References


- Boyle, D. & Harris, M. (2009). *The challenge of co-production.* London: NESTA.
- Brown, T. & Wyatt, J. (2010). Design thinking for social innovation. *Stanford Social Innovation Review*, Winter, 30-35.
- Burns, C., Cottam, H., Vanstone, C., & Winhall, J. (2006). *Transformation Design*. London: Design Council.
- Christensen, C., Bauman, H., Ruggles, R., & Sadtler, T. (2006). Disruptive innovation for social change. *Harvard Business Review*, December, 94-101.
- Christensen, C., Grossman, J., & Hwang, J. (2009). *The innovator's prescription: A disruptive solution for health care*. New York, NY: McGraw-Hill.

- Christensen, C., Horn, M., & Johnson, C. (2008). *Disrupting class: How disruptive innovation will change the way the world learns*. New York, NY: McGraw-Hill.
- Cottam, H. & Leadbeater, C. (2004). *Health: Co-creating services*. London: Design Council.
- Cross, N. (2006). *Designerly ways of knowing*. London: Springer-Verlag.
- Dong, A. (2008). The policy of design: A capabilities approach. *Design Issues*, 24 (4), 76-87.
- Immelt, J. R., Govindarajan, V., Trimble, C. (2009). How GE is disrupting itself. *Harvard Business Review,* September, 3-11.
- Lawson, B. (2006). *How designers think: The design process de-mystified* (4th ed.). Amsterdam: Elsevier.
- Muller, M. J. (2002). Participatory Design: The third space in HCI. In J. Jacko & A. Sears (Eds.), *The human-computer interaction handbook* (1051-1068). Mahwah, NJ: Lawrence Erlbaum Associates.
- Nussbaum, M. C. (2011). *Creating capabilities: The human development approach*. Cambridge, MA: Harvard University Press.
- Oosterlaken, I. (2009). Design for development. *Design Issues*, 25 (4), 91-102.
- Pals, N., Steen, M., Langley, D., & Kort, J. (2008). Three approaches to take the user perspective into account during new product design. *International Journal of Innovation Management*, 12 (3), 275-294.
- Papanek, V. (1991). *Design for the real world* (2nd ed.). London: Thames & Hudson.

- Parker, S. & Heapy, J. (2006). *The journey to the interface: How public service design can connect users to reform*. London: DEMOS.
- Prahalad, C. K. (2004). *The fortune at the bottom of the pyramid: Eradicating poverty through profits*. Warthon School Publishing.
- Sanders, E. B. N. & Stappers, P. J. (2008). Co-creation and the new landscapes of design. *CoDesign*, 4 (1), 5-18.
- Schuler, D. & Namioka, A. (Eds.) (1993). *Participatory design: Principles and practices*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Sen, A. (1999). *Development as freedom*. New York: Knopf.
- Steen, M. (2008). *The fragility of human-centred design* [PhD thesis]. Delft University of Technology.
- Steen, M. (2011a). Reflexive practice and human-centred design. *Zoontechnica*, 1 (1).
- Steen, M. (2011b). Tensions in human-centred design. *CoDesign*, 7 (1), 45-60.
- Steen, M. (2012). Human-centred design as a fragile encounter. *Design Issues*, 28 (1), 72-80.
- Steen, M., Buijs, J., & Williams, D. (forthcoming). The role of scenarios and demonstrators in promoting shared understanding in innovation projects. *International Journal of Innovation and Technology Management*.
- Steen, M., Manschot, M., & de Koning, N. (2011). Benefits of co-design in service design projects. *International Journal of Design*, 5 (2), 53-60.
- Thomas, E. (Ed.) (2008). *Innovation by design in public services*. London: Solace Foundation Imprint.

### About Marc Steen

*Marc Steen works as a senior scientist at Dutch research and innovation organization TNO. His expertise is in human-centred design (Steen, 2008; Steen, 2011a; Steen, 2011b; Steen, 2012), co-design and service design (Steen et al., 2011), and innovation management (Pals et al., 2008; Steen et al., forthcoming).*

## 17.10 COMMENTARY BY PAUL HEKKERT

### How to cite this commentary in your report

### Paul Hekkert

© *Paul Hekkert*

Paul Hekkert is professor of Form Theory at the department of Industrial Design of Delft University of Technology. His main research interest is product experience, including product aesthetics, emotion, expressiveness, and attachment. Next, he is involved in design methodology and has co-developed an interaction-centred design approach, called ViP (Vision in Product design)....

Paul Hekkert
Paul Hekkert is a member of The Interaction Design Foundation

Designers talk a lot about innovation and every designer wants to be an innovator. Well, most... Christensen sketches a few trajectories along which a company can innovate and this is a great analysis for business people, managers, and decision makers, because he meticulously explains the conditions under which a disruptive innovation might fail and succeed. Since most companies (luckily) rely more and more on designers to drive this innovation process, he also shows designers where and how they can make a difference.  If you read carefully.

Designers who just have a superficial knowledge of the innovation literature may be a bit puzzled. They may be familiar with the distinction between *incremental* – making existing stuff a little better (we will come back to this 'better') – and *radical* or *breakthrough* – coming up with a completely new product type or category – innovation. Christensen proposes a completely different distinction between sustaining and disruptive innovations. Both can be incremental and both can be radical, it all depends on their effect on the established businesses and mainstream markets. And Christensen talks about companies, technology and the absorption capacity of customers. Design, or the designer, is non-existent in his scheme of things. Let's see if we can fit him or her in somewhere.

To explain how innovations can disrupt markets, Christensen contrasts these disruptions with *sustaining* innovations. So-called high-end customers continuously demand "better performance than what was previously available". Do they? Clearly, designers can make products perform better: faster, more reliable, more sustainable, more efficient, more user-friendly, etc. This can be a small step (incremental) or a big step (breakthrough). Key is that the innovation sustains the current market and the better product can be sold with higher margins. A nice example is the (Dutch!) Senz umbrella. In a market that seemed completely satisfied, the designers came up with a radically new umbrella concept, an umbrella that is stormproof, requires a different interaction (pull instead of push to open), and looks and feels different.


Figure 17.1: The SENZ umbrella, developed at the Delft University of Technology, withstanding stormy winds. Picture taken at the Kunsthal exhibition on Dutch Design.



When it comes to disruptive innovations, Christensen makes an interesting distinction between *low-end disruptions* and *new-market disruptions*. Low-end disruptions, to begin with, typically involve cheaper, simpler, and more convenient alternatives to existing products. A nice and relatively recent example from his chart is Amazon.com (and equivalent internet bookstores) that slowly and gradually disrupt traditional bookshops. Although I know of designers who take pride in coming up with clever, cheap and low-fi alternatives for available products, it is not the kind of innovation many designers would strive for. These alternatives are foremost attractive from a business perspective.

It is the new-market disruptions that appeal to most designers – at least those that strive to be innovative. These innovations initially address new markets and new customers and gradually invade (and disrupt) the mainstream market. I however wonder whether (a low) price and ease of use, as Christensen claims, are also the only and defining qualities here. Take for example the highly acclaimed iPad. It certainly started to attract many new customers – well, who was *new* in this market? – and gradually started to compete with laptops and notebooks. But was this because it is a cheap or easy to use product? Maybe so, but foremost the iPad allowed for a completely new and compelling way of interacting with media and information. There is much more to (using the) iPad than price, functionality, and convenience; the iPad offers us a new experience, one that is compelling, engaging, cursory, and fun. Not only for individuals, but also for families!


FIGURE 17.2: The iPad offers us a new experience, one that is compelling, engaging, cursory, and fun.

*Courtesy of Minvinkallare. Copyright: pd (Public Domain (information that is common property and contains no original authorship)).*

In his excellent Chapter on Experience Design, Marc Hassenzahl argues that this is what designers can and should do: define and create new experiences. For this, as I indicated in my commentary to his chapter, and by referring to books by Verganti (2009) and myself (Hekkert & van Dijk, 2011), designers should not follow a demand from the market. Rather, they should push new markets by offering new meanings, new values, in ways that people never imagined would be possible. Interactive technologies and new media are the carriers par excellence to embody these new meanings. More and more companies start to acknowledge the power of design to drive new-market innovations that may turn out to be disruptive.

### 17.10.1 References


- Hassenzahl, Marc (2011). User Experience and Experience Design. In: Soegaard, Mads and Dam, Rikke Friis (eds.), Encyclopedia of Human-Computer Interaction.
- Hekkert, P. & van Dijk, M. (2011). Vision in Design: A Guidebook for Innovators. Amsterdam: BIS.
- Verganti, R. (2009). Design Driven Innovation – Changing the Rules of Competition by Radically Innovating what Things Mean. Boston, MA: Harvard Business Press.


## 17.11 REFERENCES


Christensen, Clayton M. (1997): *The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail*. Harvard Business Press

## ABOUT THE AUTHOR

### Clayton M. Christensen

*© Clayton M. Christensen*

Clayton M. Christensen is professor at Harvard Business School and a New York Times bestseller. He is the architect of, and the world's foremost authority on, disruptive innovation. Consistently acknowledged in rankings and surveys as one of the world's leading thinkers on innovation, his research has been applied to national economies, start-up and *Fortune 50* companies, as well as to early and late stage investing.

His seminal book *The Innovator's Dilemma* (1997), which first outlined his disruptive innovation frameworks, received the Global Business Book Award for the Best Business Book of the Year in 1997, was a *New York Tim*es bestseller, has been translated into over 10 languages, and is sold in over 25 countries. He is also a four-time recipient of the McKinsey Award for the *Harvard Business Reviews*'s best article and received a Lifetime Achievement Award from the Tribeca Film Festival in 2010.

Christensen has recently focused his innovation lens on two of our most vexing social issues, education and health care. *Disrupting Class* which looks at the root causes of why schools struggle and offers solutions was named one of the

"10 Best Innovation and Design Books in 2008" by *BusinessWeek* and the best Human Capital book of the year in the *Strategy + Business* Best Books of 2008. *The Innovator's Prescription* (2009) examines how to fix the problems facing healthcare. So as to further examine and apply his frameworks to the social sector, Christensen founded Innosight Institute, a non-profit think tank, in 2008.

An advisor to numerous countries and companies, including the government of Singapore, he is currently a board member at India's Tata Consultancy Services (NYSE: TCS), Franklin Covey (NYSE: FC), W.R. Hambrecht, and Vanu. Christensen also applies his frameworks via management consultancy Innosight which he co-founded in 2000, and Rose Park Advisors, an investment firm he founded in 2007.

Christensen was born in Salt Lake City, Utah in 1952. He graduated with highest honors in economics from Brigham Young University in 1975. Later, he received an M.Phil. in applied econometrics and the economics of less-developed countries from Oxford University in 1977, where he studied as a Rhodes Scholar. He received an MBA with High Distinction from the Harvard Business School in 1979, graduating as a George F. Baker Scholar. In 1982-1983 he was a White House fellow, serving as an assistant to U.S. Transportation Secretaries Drew Lewis and Elizabeth Dole. In 1992, he was awarded a DBA from the Harvard Business School, receiving the Best Dissertation Award from the Institute of Management Sciences for his doctoral thesis on technology development in the disk drive industry. He is currently the Robert and Jane Cizik Professor of Business Administration at the Harvard Business School.

Professor Christensen is committed to both community and church. In addition to his stint as a White House Fellow, he was an elected member of the Belmont Town Council for 8 years, and has served the Boy Scouts of America for 25 years as a scoutmaster, cub master, den leader and troop and pack committee chairman. He also worked as a missionary for the Church of Jesus Christ of Latter-day Saints in the Republic of Korea from 1971 to 1973, speaks fluent Korean, and is currently a leader in his church. He and his wife Christine live in Belmont, MA. They are the parents of five children, and have three grandchildren.

## Your notes and thoughts on chapter 17

Record your notes and thoughts on this chapter. If you want to share these thoughts with others online, go to the bottom of the page at:
http://www.interaction-design.org/encyclopedia/disruptive_innovation.html

### NOTES

# Chapter 18

# Open User Innovation

*by Eric von Hippel.*

Almost 30 years ago, researchers began a systematic study of innovation by end users and user firms. At that time, the phenomenon was generally regarded as a minor oddity. Today, it is clear that innovation by users, generally openly shared, is a very powerful and general phenomenon. It is rapidly growing due to continuing advances in computing and communication technologies. It is becoming both an important rival to and an important feedstock for producer-centered innovation in many fields. In this chapter, I provide an overview of what the international research community now understands about this phenomenon.

## 18.1 INTRODUCTION

Ever since Schumpeter (1934) promulgated his theory of economic development, economists, policymakers and business managers have assumed that the dominant mode of innovation is a "producers model." That is, it has been assumed that most important innovations would originate from producers and be supplied to consumers via goods that were for sale.

This view seemed reasonable on the face of it – producers generally serve many users and so can profit from multiple copies of a single innovative design. Individual users in contrast, depend upon benefits from in-house use of an innovation to recoup their investments. Presumably, therefore, a producer who serves many customers can afford to invest more in innovation than any single user. From this it follows logically that producer-developed designs should dominate user-developed designs in most parts of the economy.

However, the producers' model is only one mode of innovation. A second, increasingly important model is *open user innovation*. Under this second model, economically important innovations are developed by users and other agents who divide up the tasks and costs of innovation development and then *freely reveal* their results. Users obtain direct use benefits from the collaborative effort. Other participants obtain diverse benefits such as enjoyment, learning, reputation, or an increased demand for complementary goods and services.

Open user innovation is an institution that competes with and, I will argue, can displace producer innovation in many parts of the economy. A growing body of empirical work clearly shows that users are the first to develop many and perhaps most new industrial and consumer products. In addition, the importance of product and service development by users is increasing over time. This shift is being driven by two related technical trends: (1) the steadily improving *design capabilities* (innovation toolkits) that advances in computer hardware and software make possible for users; (2) the steadily improving ability of individual users to *combine and coordinate* their innovation-related efforts via new communication media such as the Internet.

The ongoing shift of innovation to users has some very attractive qualities. It is becoming progressively easier for many users to get precisely what they want by designing it for themselves. Innovation by users also provides a very necessary complement to and feedstock for manufacturer innovation. And innovation by users appears to increase social welfare. At the same time, the ongoing shift of product-development activities from manufacturers to users is painful and difficult for many manufacturers. Open, distributed innovation is "attacking" a major structure of the social division of labor. Many firms and industries must make fundamental changes to long-held business models in order to adapt. Further, governmental policy and legislation sometimes preferentially supports innovation by manufacturers. Considerations of social welfare suggest that this must change. The workings of the intellectual property system are of special concern. But despite the difficulties, a user-centered system of innovation appears well worth striving for.

Today a number of innovation process researchers are working to develop our understanding of open user innovation processes. In this chapter, I offer a review of some collective learnings on this important topic to date.

## 18.2 IMPORTANCE OF INNOVATION BY USERS

Users, as I use the term, are firms or individual consumers that expect to benefit from *using* a product or a service. In contrast, manufacturers expect to benefit from *selling* a product or a service. A firm or an individual can have different relationships to different products or innovations. For example, Boeing is a manufacturer of airplanes, but it is also a user of machine tools. If one were examining innovations developed by Boeing for the airplanes it sells, Boeing would be a manufacturer-innovator in those cases. But if one were considering innovations in metal-forming machinery developed by Boeing for in-house use in building airplanes, those would be categorized as user-developed innovations and Boeing would be a user-innovator in those cases.

Innovation user and innovation manufacturer are the two general "functional" relationships between innovator and innovation. Users are unique in that they alone benefit *directly* from innovations. All others (here lumped under the term “manufacturers”) must sell innovation-related products or services to users, indirectly or directly, in order to profit from innovations. Thus, in order to profit, inventors must sell or license knowledge related to innovations, and manufacturers must sell products or services incorporating innovations. Similarly, suppliers of innovation-related materials or services — unless they have direct use for the innovations — must sell the materials or services in order to profit from the innovations.

The user and manufacturer categorization of relationships between innovator and innovation can be extended to specific function, attributes, or features of products and services. When this is done, it may turn out that different parties are associated with different attributes of a particular product or service. For example, householders are the users of the switching attribute of a household electric light switch — they use it to turn lights on and off. However, switches also have other attributes, such as “easy wiring” qualities, that may be used only by the electricians who install them. Therefore, if an electrician were to develop an improvement to the installation attributes of a switch, it would be considered a user-developed innovation.

Both qualitative observations and quantitative research in a number of fields clearly document the important role users play as first developers of products and services later sold by manufacturing firms. Adam Smith (1776) was an early observer of the phenomenon, pointing out the importance of “the invention of a great number of machines which facilitate and abridge labor, and enable one man to do the work of many.” Smith went on to note that “a great part of the machines made use of in those manufactures in which labor is most subdivided, were originally the invention of common workmen, who, being each of them employed in some very simple operation, naturally turned their thoughts towards finding out easier and readier methods of performing it.” Rosenberg (1976) explored the matter in terms of innovation by *user firms* rather than individual workers. He studied the history of the US machine tool industry, finding that important and basic machine types like lathes and milling machines were first developed and built by user firms having a strong need for them. Textile manufacturing firms, gun manufacturers and sewing machine manufacturers were important early user-developers of machine tools.

Quantitative studies of user innovation document that many of the most important and novel products and processes in a range of fields have been developed by user firms and by individual users. Thus, Enos (1962) reported that nearly all the most important innovations in oil refining were developed by user firms. Freeman (1968) found that the most widely licensed chemical production processes were developed by user firms. Von Hippel (1988) found that users were the developers of about 80 percent of the most important scientific instrument innovations, and also the developers of most of the major innovations in semiconductor processing. Pavitt (1984) found that a considerable fraction of invention by British firms was for in-house use. Shah (2000) found that the most commercially important equipment innovations in four sporting fields tended to be developed by individual users.

Empirical studies also show that *many* users — from 10 percent to nearly 40 percent — engage in developing or modifying products. This has been documented in the case of specific types of industrial products and consumer products, and in large, multi-industry studies of process innovation in Canada and the Netherlands as well (table 1). When taken together, the findings make it very clear that users are doing a *lot* of product development and product modification in many fields.

| Innovation Area | Number and type of users sampled | % developing and building product for own use |
|---|---|---|
| **Industrial products** | | |
| 1. Printed Circuit CAD Software (a) | 136 user firm attendees at a PC-CAD conference | 24.3% |
| 2. Pipe Hanger Hardware (b) | Employees in 74 pipe hanger installation firms | 36% |
| 3. Library Information Systems (c) | Employees in 102 Australian libraries using computerized OPAC library information systems | 26% |
| 4. Medical Surgery Equipment (d) | 261 surgeons working in university clinics in Germany | 22% |
| 5. Apache OS server software security features (e) | 131 technically sophisticated Apache users (webmasters) | 19.1% |
| **Consumer products** | | |
| 6. Outdoor consumer products (f) | 153 recipients of mail order catalogs for outdoor activity products for consumers | 9.8% |
| 7. "Extreme" sporting equipment (g) | 197 members of 4 specialized sporting clubs in 4 "extreme" sports | 37.8% |

| | | |
|---|---|---|
| 8. Mountain biking equipment (h) | 291 mountain bikers in a geographic region known to be an "innovation hot spot." | 19.2% |
| **Multi-industry process innovation surveys** | | |
| 26 'Advanced Manufacturing Technologies' (i) | Canadian manufacturing plants in 9 Manufacturing Sectors (less food processing) in Canada, 1998 (population estimates based upon a sample of 4,200) | 28% developed<br>26% modified |
| 39 'Advanced Manufacturing Technologies' (j) | 16,590 Canadian manufacturing establishments that met the criteria of having at least $250,000 in revenues, and at least 20 employees. | 22% developed<br>21% modified |
| Any type of process innovation or process modification (k) | Representative, cross-industry sample of 498 "high tech" Netherlands SMEs | 41% developed only 34% modified only 54% developed and/or modified |

Table 18.1: Studies of user innovation frequency. Sources of Data: (a) Urban and von Hippel (1988); (b) Herstatt and von Hippel (1992); (c) Morrison et al. (2000); (d) Lüthje (2003); (e) Franke and von Hippel (2003); (f) Lüthje (2004); (g) Franke and Shah (2003); (h) Lüthje et al. (2002); (i) Arundel and Sonntag 1999; (j) Gault and von Hippel 2009; (k) de Jong and von Hippel 2009.

Studies of innovating users (both individuals and firms) show them to have the characteristics of "lead users" (Urban & von Hippel 1988, Herstatt and von Hippel 1992, Olson and Bakke 2001, Lilien et al. 2002). That is, they are ahead of the majority of users in their populations with respect to an important market trend, and they expect to gain relatively high benefits from a solution to the needs they have encountered there. The correlations found between innovation by users and lead user status are highly significant, and the effects are very large (Franke & Shah 2003, Lüthje et al. 2002 and Morrison et al. 2000).

Since lead users are at the leading edge of the market with respect to important market trends, one can guess that many of the novel products they develop for their own use will appeal to other users too and so might provide the basis for products manufacturers would wish to commercialize. This turns out to be the case. A number of studies have shown that many of the innovations reported by lead users are judged to be commercially attractive and/or have actually been commercialized by manufacturers.

Research provides a firm grounding for these empirical findings. The two defining characteristics of lead users and the likelihood that they will develop new or modified products have been found to be highly correlated (Morrison et al. 2004). In addition, it has been found that the higher the intensity of lead user characteristics displayed by an innovator, the greater the commercial attractiveness of the innovation that that lead user develops (Franke and von Hippel 2003a). In Figure 18.1, the increased concentration of innovations toward the right indicates that the likelihood of innovating is higher for users having higher lead user index values. The rise in average innovation attractiveness as one moves from left to right indicates that innovations developed by lead users tend to be more commercially attractive. (Innovation attractiveness is the sum of the novelty of the innovation and the expected future generality of market demand.)

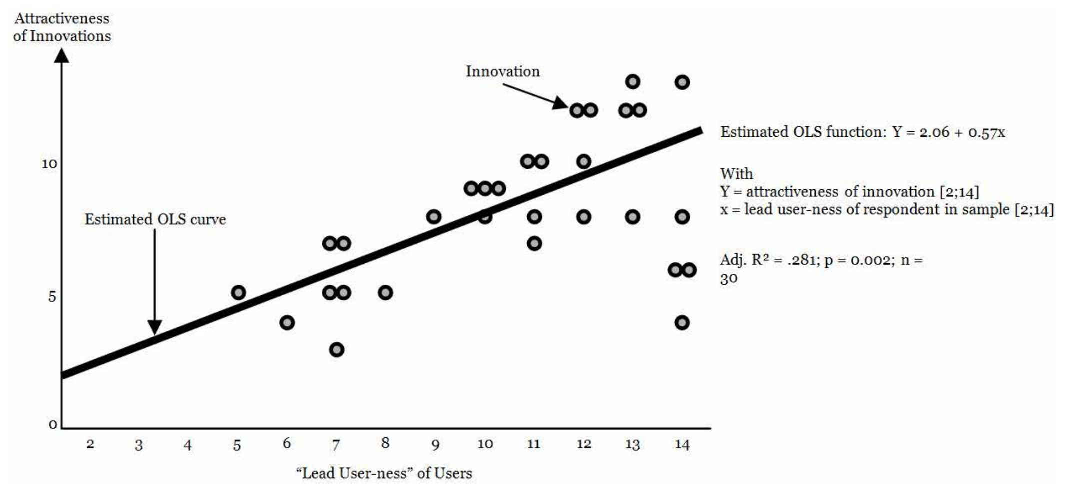



Figure 18.1: User-innovators with stronger "lead user" characteristics develop innovations having higher appeal in the general marketplace. Data Source: Franke and von Hippel 2003.



## 18.3 WHY MANY USERS WANT CUSTOM PRODUCTS

Why do so many users develop or modify products for their own use? Users may innovate if and as they want something that is not available on the market and are able and willing to pay for its development. It is likely that many users do not find what they want on the market. Meta-analysis of market-segmentation studies suggests that users' needs for products are highly heterogeneous in many fields (Franke and Reisinger 2003).

Mass producers tend to follow a strategy of developing products that are designed to meet the needs of a large market segment well enough to induce purchase from and capture significant profits from a large number of customers. When users' needs are heterogeneous, this strategy of "a few sizes fit all" will leave many users somewhat dissatisfied with the commercial products on offer and probably will leave some users seriously dissatisfied. In a study of a sample of users of the security features of Apache web server software, Franke and von Hippel (2003b) found that users had a very high heterogeneity of need, and that many had a high willingness to pay to get precisely what they wanted. Nineteen percent of the users sampled actually innovated to tailor Apache more closely to their needs. Those who did were found to be significantly more satisfied.

## 18.4 USERS' INNOVATE-OR-BUY DECISIONS

Even if many users want "exactly right products" and are willing and able to pay for their development, we must understand why users often do this for themselves rather than hire a custom producer to develop a special just-right product for them. After all, custom producers specialize in developing products for one or a few users. Since these firms are specialists, it is possible that they could design and build custom products for individual users or user firms faster, better, or cheaper than users could do this for themselves. Despite this possibility, several factors can drive users to innovate rather than buy. Both in the case of user firms and in the case of individual user-innovators, agency costs play a major role. In the case of individual user-innovators, enjoyment of the innovation process can also be important.

With respect to agency costs, consider that when a user develops its own custom product that user can be trusted to act in its own best interests. When a user hires a producer to develop a custom product, the situation is more complex. The user is then a principal that has hired the custom producer to act as its agent. If the interests of the principal and the agent are not the same, there will be agency costs. In general terms, agency costs are (1) costs incurred to monitor the agent to ensure that it (or he or she) follows the interests of the principal, (2) the cost incurred by the agent to commit itself not to act against the principal's interest (the "bonding cost"), and (3) costs associated with an outcome that does not fully serve the interests of the principal (Jensen and Meckling 1976). In the specific instance of product and service development, a major divergence of interests between user and custom producer does exist: the user wants to get precisely what it needs, to the extent that it can afford to do so. In contrast, the custom producer wants to lower its development costs by incorporating solution elements it already has or that it predicts others will want in the future — even if by doing so it does not serve its present client's needs as well as it could.

A user wants to preserve its need specification because that specification is chosen to make *that user's* overall solution quality as high as possible at the desired price. For example, an individual user may specify a mountain-climbing boot that will precisely fit his unique climbing technique and allow him to climb Everest more easily. Any deviations in boot design will require compensating modifications in the climber's carefully practiced and deeply ingrained climbing technique — a much more costly solution from the user's point of view. A custom boot producer, in contrast, will have a strong incentive to incorporate the materials and processes it has in stock and expects to use in future even if this produces a boot that is not precisely right for the present customer. For example, the producer will not want to learn a new way to bond boot components together even if that would produce the best custom result for one client. The net result is that when one or a few users want something special they will often get the best result by innovating for themselves.

A model of the innovate-or-buy decision (von Hippel 2005) shows in a quantitative way that user firms with unique needs (in other words, a market of one) will always be better off developing new products for themselves. It also shows that development by producers can be the most economical option when $n$ or more user firms want the same thing. However, when the number of user firms wanting the same thing lies between 1 and $n$, producers may not find it profitable to develop a new product for just a few users. In that case, more than one user may invest in developing the same thing independently, owing to market failure.

This results in a waste of resources from the point of view of social welfare. The problem can be addressed by new institutional forms, such as the user innovation communities that will be mentioned later.

It is important to note that an additional incentive can drive individual user-innovators to innovate rather than buy: they may value the *process* of innovating because of the enjoyment or learning that it brings them. It might seem strange that user-innovators can enjoy product development enough to want to do it themselves — after all, producers pay their product developers to do such work! On the other hand, it is also clear that enjoyment of problem solving is a motivator for many individual problem solvers in at least some fields. Consider for example the millions of crossword-puzzle aficionados. Clearly, for these individuals enjoyment of the problem-solving process rather than the solution is the goal. One can easily test this by attempting to offer a puzzle solver a completed puzzle — the very output he or she is working so hard to create. One will very likely be rejected with the rebuke that one should not spoil the fun. Pleasure as a motivator can apply to the development of commercially useful innovations as well. Studies of the motivations of volunteer contributors of code to widely used software products have shown that these individuals too are often strongly motivated to innovate by the joy and learning they find in this work (Hertel et al. 2003; Lakhani and Wolf 2005).

## 18.5 USERS' LOW-COST INNOVATION NICHES

An exploration of the basic processes of product and service development show that users and producers tend to develop different *types* of innovations. This is due in part to information asymmetries: users and producers tend to know different things. Product developers need two types of information in order to succeed at their work: need and context-of-use information (generated by users) and generic solution information (often initially generated by producers specializing in a particular type of solution). Bringing these two types of information together is not easy. Both need information and solution information are often very "sticky" — that is, costly to move from the site where the information was generated to other sites (von Hippel 1994). It should be noted that observation that information is often sticky contravenes a central tendency in economic theorizing. Much of the research on the special character of markets for information and the difficulty of appropriating benefit from invention and innovation has been based on the idea that information can be transferred at very low cost. Thus, Arrow (1962) observes that "the cost of transmitting a given body of information is frequently very low. . . . In the absence of special legal protection, the owner cannot, however, simply sell information on the open market. Any one purchaser can destroy the monopoly, since he can reproduce the information at little or no cost" (Arrow 1962, 614-15).

When information is sticky, innovators tend to rely largely on information they already have in stock. One consequence of the resulting typical asymmetry between users and producers is that users tend to develop innovations that are functionally novel, requiring a great deal of user-need information and use-context information for their development. In contrast, producers tend to develop innovations that are improvements on well-known needs and that require a rich understanding of solution information for their development. Similarly, users tend to have better information regarding ways to improve use-related activities such as maintenance than do producers: they "learn by using" (Rosenberg 1982).

This sticky information effect is quantitatively visible in studies of innovation. Riggs and von Hippel (1994) studied the types of innovations made by users and producers that improved the functioning of two major types of scientific instruments. They found that users are significantly more likely than producers to develop innovations that enabled the instruments to do qualitatively new types of things for the first time. In contrast, producers tended to develop innovations that enabled users to do the same things they had been doing, but to do them more conveniently or reliably (table 2). For example, users were the first to modify the instruments to enable them to image and analyze magnetic domains at submicroscopic dimensions. In contrast, producers were the first to computerize instrument adjustments to improve ease of operation. Sensitivity, resolution, and accuracy improvements fall somewhere in the middle, as the data show. These types of improvements can be driven by users seeking to do specific new things, or by producers applying their technical expertise to improve the products along known general dimensions of merit, such as accuracy.

| Type of improvement provided by innovation | | Innovation developed by: | | | |
|---|---|---|---|---|---|
| | | %User | User | Producer | Total |
| (1) New functional capability | | 82% | 14 | 3 | 17 |
| (2) Sensitivity, resolution, or accuracy improvement | | 48% | 11 | 12 | 23 |
| (3) Convenience or reliability improvement | | 13% | 3 | 21 | 24 |
| | | | | Total | 64 |

Table 18.2: Source of innovations by nature of improvement effected. Source: Riggs and von Hippel (1994).

The sticky information effect is independent of Stigler's (argument that the division of labor is limited by the extent of the market (Stigler 1951). When profit expectations are controlled for, the impact of sticky information on the locus of innovation is still strongly evident (Ogawa 1998).

If we extend the information-asymmetry argument one step further, we see that information stickiness implies that information on hand will also differ among *individual* users and producers. The information assets of some particular user (or some particular producer) will be closest to what is required to develop a particular innovation, and so the cost of developing that innovation will be relatively low for that user or producer. The net result is that user innovation activities will be *distributed* across many users according to their information endowments. With respect to innovation, one user is by no means a perfect substitute for another.

## 18.6 WHY USERS OFTEN FREELY REVEAL THEIR INNOVATIONS

The social efficiency of a system in which individual innovations are developed by individual users is increased if users somehow diffuse what they have developed to others. Producer-innovators *partially* achieve this when they sell a product or a service on the open market (partially because they diffuse the product incorporating the innovation, but often not all the information that others would need to fully understand and replicate it). If user-innovators do not somehow also diffuse what they have done, multiple users with very similar needs will have to independently develop very similar innovations — a poor use of resources from the viewpoint of social welfare. Empirical research shows that users often do achieve widespread diffusion by an unexpected means: they often "freely reveal" what they have developed. When we say that an innovator freely reveals information about a product or service it has developed, we mean that all intellectual property rights to that information are voluntarily given up by the innovator, and all interested parties are given access to it — the information becomes a public good (Harhoff et al 2003).

The empirical finding that users often freely reveal their innovations has been a major surprise to innovation researchers. On the face of it, if a user-innovator's proprietary information has value to others, one would think that the user would strive to prevent free diffusion rather than help others to free ride on what it has developed at private cost. Nonetheless, it is now very clear that individual users and user firms — and sometimes producers — often freely reveal detailed information about their innovations.

The practices visible in "open source" software development were important in bringing this phenomenon to general awareness. In these projects it was clear *policy* that project contributors would routinely and systematically freely reveal code they had developed at private expense (Raymond 1999). However, free revealing of product innovations has a history that began long before the advent of open source software. Allen, in his 1983 study of the eighteenth-century iron industry, was probably the first to consider the phenomenon systematically (Allen 1983). Later, Nuvolari (2004) discussed free revealing in the early history of mine pumping engines. Contemporary free revealing by users has been documented by von Hippel and Finkelstein (1979) for medical equipment, by Lim (2000) for semiconductor process equipment, by Morrison et al (2000) for library information systems, and by Franke and Shah (2003) for sporting equipment. Henkel (2003) has documented free revealing among producers in the case of embedded Linux software.

Innovators often freely reveal because it is often the best or the only practical option available to them. Hiding an innovation as a trade secret is unlikely to be successful for long: too many generally know similar things, and some holders of the "secret" information stand to lose little or nothing by freely revealing what they know. Studies find that innovators in many fields view patents as having only limited value (Harhoff et al, 2003). Copyright protection and copyright licensing are applicable only to "writings," such as books, graphic images, and computer software.

Active efforts by innovators to freely reveal — as opposed to sullen acceptance — are explicable because free revealing can provide innovators with significant private benefits as well as losses or risks of loss. Users who freely reveal what they have done often find that others then improve or suggest improvements to the innovation, to mutual benefit (Raymond 1999). Freely revealing users also may benefit from enhancement of reputation, from positive network effects due to increased diffusion of their innovation, and from other factors. Being the first to freely reveal a particular innovation can also enhance the benefits received, and so there can actually be a rush to reveal, much as scientists rush to publish in order to gain the benefits associated with being the first to have made a particular advancement.

## 18.7 INNOVATION COMMUNITIES

Innovation by users tends to be widely distributed rather than concentrated among just a very few very innovative users (table 3). As a result, it is important for user-innovators to find ways to combine and leverage their efforts. Users achieve this by engaging in many forms of cooperation. Direct, informal user-to-user cooperation (assisting others to innovate, answering questions, and so on) is common. Organized cooperation is also common, with users joining together in networks and communities that provide useful structures and tools for their interactions and for the distribution of innovations. Innovation communities can increase the speed and effectiveness with which users and also producers can develop and test and diffuse their innovations. They also can greatly increase the ease with which innovators can build larger systems from interlinkable modules created by community participants.

| **User samples** | **Number of innovations each user developed:** | | | | | |
|---|---|---|---|---|---|---|
| | **1** | **2** | **3** | **6** | **na** | **sample (n)** |
| Scientific Instrument users* | 28 | 0 | 1 | 0 | 1 | 32 |
| Scientific Instrument users** | 20 | 1 | 0 | 1 | 0 | 28 |
| Process equipment users*** | 19 | 1 | 0 | 0 | 8 | 29 |
| Sports equipment users**** | 7 | 0 | 0 | 0 | 0 | 7 |
| | | | | | | |

Table 18.3: User innovation is widely distributed: Few users developed more than one major commercialized innovation. Table Source: von Hippel (2005), table 7-1. Data Sources: * von Hippel 1988, Appendix: GC, TEM, NMR Innovations; ** Riggs and von Hippel, Esca and AES; *** von Hippel 1988, Appendix: Semiconductor and pultrusion process equipment innovations; **** Shah 2000, Appendix A: skateboarding, snowboarding and windsurfing innovations developed by users.

Free and open source software projects are a relatively well-developed and very successful form of Internet-based innovation community. However, innovation communities are by no means restricted to software or even to information products, and they can play a major role in the development of physical products. Franke and Shah (2003) have documented the value that user innovation communities can provide to user-innovators developing physical products in the field of sporting equipment. The analogy to open source innovation communities is clear.

The collective or community effort to provide a public good — which is what freely revealed innovations are — has traditionally been explored in the literature on “collective action.” However, behaviors seen in extant innovation communities fail to correspond to that literature at major points. In essence, innovation communities appear to be more robust with respect to recruiting and rewarding members than the literature would predict. The reason for this appears to be that innovation contributors obtain some private rewards that are not shared equally by free riders (those who take without contributing). For example, a product that a user-innovator develops and freely reveals might be perfectly suited to that user-innovator's requirements but less well suited to the requirements of free riders. Innovation communities thus illustrate a "private-collective" model of innovation incentive (von Hippel and von Krogh 2003).

## 18.8 ADAPTING POLICY TO USER INNOVATION

Is innovation by users a "good thing?" Welfare economists answer such a question by studying how a phenomenon or a change affects social welfare. Henkel and von Hippel (2005) explored the social welfare implications of user innovation. They found that, relative to a world in which only producers innovate, social welfare is very probably increased by the presence of innovations freely revealed by users. This finding implies that policy making should support user innovation, or at least should ensure that legislation and regulations do not favor producers at the expense of user-innovators.

The transitions required of policy making to achieve neutrality with respect to user innovation vs. producer innovation are significant. Consider the impact on open and distributed innovation of past and current policy decisions. Research done in the past 30 years has convinced many academics that intellectual property law is sometimes or often not having its intended effect. Intellectual property law was intended to increase the amount of innovation investment. Instead, it now appears that there are economies of scope in both patenting and copyright that allow firms to use these forms of intellectual property law in ways that are directly opposed to the intent of policy makers and to the public welfare (Foray 2004). Major firms can invest to develop large portfolios of patents. They can then use these to create "patent thickets" — dense networks of patent claims that give them plausible grounds for threatening to sue across a wide range of intellectual property. They may do this to prevent others from introducing a superior innovation and/or to demand licenses from weaker competitors on favorable terms (Shapiro 2001, Bessen 2003). Movie, publishing, and software firms can use large collections of copyrighted work to a similar purpose (Benkler 1999). In view of the distributed nature of innovation by users, with each tending to create a relatively small amount of intellectual property, users are likely to be disadvantaged by such strategies.

It is also important to note that users (and producers) tend to build prototypes of their innovations economically by modifying products already available on the market to serve a new purpose. Laws such as the (U.S.) Digital Millennium Copyright Act, intended to prevent consumers from illegally copying protected works, also can have the unintended side effect of preventing users from modifying products that they purchase (Varian 2002). Both fairness and social welfare considerations suggest that innovation-related policies should be made neutral with respect to the sources of innovation.

It may be that current impediments to user innovation will be solved by legislation or by policy making. However, beneficiaries of existing law and policy will predictably resist change. Fortunately, a way to get around some of these problems is in the hands of innovators themselves. Suppose many innovators in a particular field decide to freely reveal what they have developed, as they often have reason to do. In that case, users can collectively create an information commons (a collection of information freely available to all) containing substitutes for some or a great deal of information now held as private intellectual property. Then user-innovators can work around the strictures of intellectual property law by simply using these freely revealed substitutes (Lessig 2001).

This pattern is happening in the field of software – and very visibly so. For many problems, user-innovators in that field now have a choice between proprietary, closed software provided by Microsoft and other firms and open source software that they can legally download from the Internet and legally modify as they wish to serve their own specific needs. It is also happening, although less visibly, in the case of process equipment developed by users for in-house use. Data from both Canada and the Netherlands show that about 25% of such user-developed innovations get voluntarily transferred to producers. A significant fraction – about half – being transferred both unprotected by intellectual property and without charge (Gault and von Hippel 2009, de Jong and von Hippel 2009).

Policy making that levels the playing field between users and producers will force more rapid change onto producers but will by no means destroy them. Experience in fields where open and distributed innovation processes are far advanced show how producers can and do adapt. Some, for example, learn to supply proprietary platform products that offer user-innovators a framework upon which to develop and use their improvements (Jeppesen 2004).

## 18.9 DIFFUSION OF USER-DEVELOPED INNOVATIONS

Products, services, and processes developed by users become more valuable to society if they are somehow diffused to others that can also benefit from them. If user innovations are not diffused, multiple users with very similar needs will have to invest to (re)develop very similar innovations which, as was noted earlier, would be a poor use of resources from the social welfare point of view. In the case of information products, users have the possibility of largely or completely doing without the services of producers. Open source software projects are object lessons that teach us that users can create, produce, diffuse, provide user field support for, update, and use complex products by and for themselves in the context of user innovation communities. In physical product fields, the situation is different. Users can develop products. However, the economies of scale associated with manufacturing and distributing physical products give producers an advantage over "do-it-yourself" users in those activities.

How can or should user innovations of general interest be transferred to producers for large-scale diffusion? We propose that there are three general methods for accomplishing this. First, producers can actively seek innovations developed by lead users that can form the basis for a profitable commercial product. Second, producers can draw innovating users into joint design interactions by providing them with "toolkits for user innovation." Third, users can become producers in order to widely diffuse their innovations. We discuss each of these possibilities in turn.

To systematically find user-developed innovations, producers must redesign their product development processes. Currently, almost all producers think that their job is to find a need and fill it rather than to sometimes find and commercialize an innovation that lead users have already developed. Accordingly, producers have set up market-research departments to explore the needs of users in the target market, product-development groups to think up suitable products to address those needs, and so forth. In this type of product development system, the needs and prototype solutions of lead users — if encountered at all — are typically rejected as outliers of no interest. Indeed, when lead users' innovations do enter a firm's product line they typically arrive with a lag and by an unconventional and unsystematic route. For example, a producer may "discover" a lead user innovation only when the innovating user firm contacts the producer with a proposal to produce its design in volume to supply its own in-house needs. Or sales or service people employed by a producer may spot a promising prototype during a visit to a customer's site.

Modification of firms' innovation processes to *systematically* search for and further develop innovations created by lead users can provide producers with a better interface to the innovation process as it actually works, and so provide better performance. A natural experiment conducted at 3M illustrates this possibility. Annual sales of lead user product ideas generated by the average lead user project at 3M were conservatively forecast by management to be more than 8 times the sales forecast for new products developed in the traditional manner — $146 million versus $18 million per year. In addition, lead user projects were found to generate ideas for new product lines, while traditional market-research methods were found to produce ideas for incremental improvements to existing product lines. As a consequence, 3M divisions funding lead user project ideas experienced their highest rate of major product line generation in the past 50 years (Lilien et al. 2002).

Toolkits for user innovation custom design involve partitioning product-development and service-development projects into solution-information-intensive subtasks and need-information-intensive subtasks. Need-intensive subtasks are then assigned to users along with a kit of tools that enable them to effectively execute the tasks assigned to them. In the case of physical products, the designs that users create using a toolkit are then transferred to producers for production (von Hippel and Katz 2002). Toolkits make innovation cheaper for users and also lead to higher customer value. Thus, Franke and Piller (2004) in a study of a consumer wrist watches, found the willingness to pay for a self-designed products was 200% of the willingness to pay for the best-selling commercial product of the same technical quality. This increased willingness to pay was due to both the increased value provided by the self-developed product and the value of the toolkit process for consumers engaging in it. (Schreier and Franke 2004).

Producers that offer toolkits to their customers can attract innovating users into a relationship with their firm and so get an advantage with respect to producing what the users develop. The custom semiconductor industry was an early adopter of toolkits. In 2003, more than $15 billion worth of semiconductors were produced that had been designed using this approach. (Thomke and von Hippel 2002).

Innovations developed by users sometimes achieve widespread diffusion when those users become producers - setting up a firm to produce their innovative product(s) for sale. Shah (2000) showed this pattern in sporting goods fields. In the medical field, Lettl and Gemünden (2005) have shown a pattern in which innovating users take on many of the entrepreneurial functions needed to commercialize the new medical products they have developed, but do not themselves abandon their user roles. New work in this field is exploring the conditions under which users will become entrepreneurs rather than transfer their innovations to established firms (Hienerth 2004, Shah and Tripsas 2004).

## 18.10 SUMMARY

I summarize this overview article by again saying that users' ability to innovate is improving *radically* and *rapidly* as a result of the steadily improving quality of computer software and hardware, improved access to easy-to-use tools and components for innovation, and access to a steadily richer innovation commons. Today, user firms and even individual hobbyists have access to sophisticated programming tools for software and sophisticated CAD design tools for hardware and electronics. These information-based tools can be run on a personal computer, and they are rapidly coming down in price. As a consequence, innovation by users will continue to grow even if the degree of heterogeneity of need and willingness to invest in obtaining a precisely right product remains constant.

Equivalents of the innovation resources described above have long been available within corporations to a few. Senior designers at firms have long been supplied with engineers and designers under their direct control, and with the resources needed to quickly construct and test prototype designs. The same is true in other fields, including automotive design and clothing design: just think of the staffs of engineers and model makers supplied so that top auto designers can quickly realize and test their designs.

But if, as we have seen, the information needed to innovate in important ways is widely distributed, the traditional pattern of concentrating innovation-support resources on a few individuals is hugely inefficient. High-cost resources for innovation support cannot efficiently be allocated to "the right people with the right information:" it is very difficult to know who these people may be before they develop an innovation that turns out to have general value. When the cost of high-quality resources for design and prototyping becomes very low (the trend we have described), these resources can be diffused very widely, and the allocation problem diminishes in significance. The net result is a pattern of in which development of product and service innovations is increasingly shifting to users – a pattern that will involve significant changes for both users and producers.

## 18.11 REFERENCES


*Allen, Robert Carson (1983): Collective invention. In Journal of Economic Behavior & Organization, 4 (1) pp. 1-24*

*Arrow, Kenneth (1962): Economic Welfare and the Allocation of Resources for Invention. In: (ed.). "The Rate and Direction of Inventive Activity: Economic and Social Factors: a Conference of the Universities - National Bureau Committee for Economic Research and the Committee on Economic Growth of the Social Science Research Council". Princeton University Presspp. 609-626*

*Arundel, Anthony and Sonntag, Viki (1999). Patterns of advanced manufacturing technology (AMT) use in Canadian manufacturing: 1998 AMT survey results : final report. Science, Innovation and Electronic Information Division, Statistics Canada*

*Benkler, Yochai (1999): Intellectual property and the organization of information production. In International Review of law and Economics, 22 (1) pp. 81-107*

*Bessen, James (2002): Patent Thickets: Strategic Patenting of Complex Technologies. In Working Paper Series,*

*Enos, John L. (1962): Petroleum Progress and Profits: A History of Process Innovation. The MIT Press*

*Foray, Dominique (2004): The Economics of Knowledge. The MIT Press*

*Franke, Nikolaus and Piller, Frank (2004): Value Creation by Toolkits for User Innovation and Design: The Case of the Watch Market. In Journal of Product Innovation Management, 21 (6) pp. 401-415*

*Franke, Nikolaus and Reisinger, H. (2003). Remaining Within-Cluster Variance: a Meta-Analysis of the "Dark Side of Cluster Analysis". Working paper. Vienna Business University*

*Franke, Nikolaus and Shah, Sonali (2003): How communities support innovative activities: an exploration of assistance and sharing among end-users. In Research Policy, 32 (1) pp. 157-178*

*Freeman, Christopher (1968): Chemical Process Plant : Innovation and the World Market. In National Institute Economic Review, 45 (1) pp. 29-51*

*Harhoff, Dietmar, Henkel, Joachim and Hippel, Eric von (2003): Profiting from voluntary information spillovers: how users benefit by freely revealing their innovations. In Research Policy, 32 (10) pp. 1753-1769*

*Henkel, Joachim (2003): Software Development in Embedded Linux – Informal Collaboration of Competing Firms. In: Uhr, Wolfgang, Esswein, Werner and Schoop, Eric (eds.). "Wirtschaftsinformatik 2003 / Band II: Medien - Märkte - Mobilität (German Edition)". Physica-Verlag HDpp. 81-99*

*Henkel, Joachim and Hippel, Eric von (2004): Welfare Implications of User Innovation. In The Journal of Technology Transfer, 30 (1) pp. 73-87*

*Herstatt, Cornelius and Hippel, Eric von (1992): From Experience: Developing New Product Concepts Via the Lead User Method: A Case Study in a Low Tech Field. In Journal of Product Innovation Management, 9 (3) pp. 213-221*

*Hertel, Guido, Niedner, Sven and Herrmann, Stefanie (2003): Motivation of software developers in Open Source projects: an Internet-based survey of contributors to the Linux kernel. In Research Policy, 32 (7) pp. 1159-1177*

*Hienerth, Christoph (2004). The commercialization of user innovations: Sixteen cases in an extreme sporting industry. Vienna University of Economics and Business Administration Working Paper. Vienna University of Economics and Business Administration*

*Hippel, Eric von (1994a): "Sticky information" and the locus of problem solving: implications for innovation. In Management Science, 40 (4) pp. 429-439*

*Hippel, Eric von (2005): Democratizing Innovation. The MIT Press*

*Hippel, Eric von (1994b): The Sources of Innovation. Oxford University Press, USA*

*Hippel, Eric von and Finkelstein, Stan N. (1979): Product designs which encourage -- or discourage -- related innovation by users: an analysis of innovation in automated clinical chemistry analyzers. In Science & Public Policy, 6 (1) pp. 24-37*

*Hippel, Eric von and Katz, Ralph (2002): Shifting Innovation to Users via Toolkits. In Management Science, 48 (7) pp. 821-833*

*Hippel, Eric von and Krogh, Georg von (2003): Open Source Software and the "Private-Collective" Innovation Model: Issues for Organization Science. In Organization Science, 14 (2) pp. 209-223*

Jensen, Michael C. and Meckling, William H. (1976): *Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure. In ,*

Jong, Jeroen P. J. De and Hippel, Eric von (2009): *Measuring User Innovation in Dutch High Tech SMEs: Frequency, Nature and Transfer to Producers. In ,*

Lakhani, Karim and Wolf, Robert G. (2005): *Why Hackers Do What They Do: Understanding Motivation and Effort in Free/Open Source Software Projects. In:* Feller, Joseph, Fitzgerald, Brian, Hissam, Scott A., Lakhani, Karim R., Shirky, Clay and Cusumano, Michael *(eds.). "Perspectives on Free and Open Source Software". The MIT Press*

Lessig, Lawrence (2001): *The Future of Ideas: The Fate of the Commons in a Connected World. Random House*

Lettl, Christopher and Gemünden, Hans Georg (2005): *The entrepreneurial role of innovative users. In Journal of Business & Industrial Marketing, 20 (7) pp. 339-346*

Lilien, Gary L., Morrison, Pamela D., Searls, Kathleen, Sonnack, Mary and Hippel, Eric von *(2002): Performance Assessment of the Lead User Idea-Generation Process for New Product Development. In Management Science, 48 (8) pp. 1042-1059*

Lim, Kwanghui *(2000). The Many Faces of Absorptive Capacity: Spillovers of Copper Interconnect Technology for Semiconductor Chips. MIT Sloan School of Management. Working paper # 4110. MIT Sloan School of Management*

Lüthje, Christian *(2003): Customers as co-inventors: An empirical analysis of the antecedents of customer-driven innovations in the field of medical equipment. In: Proceedings of the 32th EMAC Conference 2003, Glasgow.*

Morrison, Pamela D., Roberts, John H. and Hippel, Eric von *(2000): Determinants of User Innovation and Innovation Sharing in a Local Market. In Management Science, 46 (12) pp. 1513-1527*

Morrison, Pamela D., Roberts, John H. and Midgley, David F. *(2004): The nature of lead users and measurement of leading edge status. In Research Policy, 33 (2) pp. 351-362*

Nuvolari, Alessandro *(2004): Collective invention during the British Industrial Revolution: the case of the Cornish pumping engine. In Cambridge Journal of Economics, 28 (3) pp. 347-363*

Ogawa, Susumu *(1997). Does sticky information affect the locus of innovation? : evidence from the Japanese convenience-store industry. Working paper (Sloan School of Management) ; WP 3984-97.*

*Ogawa, Susumu (1998): Does sticky information affect the locus of innovation? Evidence from the Japanese convenience-store industry. In Research Policy, 26 (7) pp. 777-790*

*Olson, Erik L and Bakke, Geir (2001): Implementing the lead user method in a high technology firm: A longitudinal study of intentions versus actions. In Journal of Product Innovation Management, 18 (6) pp. 388-395*

*Pavitt, Keith (1984): Sectoral patterns of technical change: Towards a taxonomy and a theory. In Research Policy, 13 (6) pp. 343-373*

*Raymond, Eric S. (1999): The Cathedral and the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary (O'Reilly Linux). OReilly Media*

*Riggs, William and Hippel, Eric von (1994): The Impact of Scientific and Commercial Values on the Sources of Scientific Instrument Innovation. In Research Policy, 23 (0) pp. 459-469*

*Rosenberg, Nathan (1976): Perspectives on Technology. Cambridge University Press*

*Rosenberg, Nathan (1982): Inside the Black Box: Technology and Economics. Cambridge University Press*

*Schreier, Martin and Franke, Nikolaus (2004). Tom Sawyer's great law in action: Why users are willing to pay to design their own products via toolkits for user innovation and design. Working Paper. Vienna University of Economics and Business Administration*

*Schumpeter, Joseph A. (1934): The Theory of Economic Development: An Inquiry into Profits, Capital, Credit, Interest, and the Business Cycle (Social Science Classics Series). Oxford University Press*

*Shah, Sonali (2000). Sources and Patterns of Innovation in a Consumer Products Field : Innovations in Sporting Equipment Sonali Shah * Massachusetts Institute of Technology Sloan Working Paper # 4105 March 2000 Initial Draft January 2000 Revised November 2000. Massachusetts Institute of Technology*

*Shah, Sonali and Tripsas, Mary (2004). When Do User-Innovators Start Firms? Towards a Theory of User Entrepreneurship. University of Illinois at Urbana-Champaign, College of Business*

*Shapiro, Carl (2001): Navigating the Patent Thicket: Cross Licenses, Patent Pools, and Standard-Setting. In: Jaffe, Adam B., Lerner, Josh and Stern, Scott (eds.). "Innovation Policy and the Economy, Vol. 1". The MIT Presspp. 119-150*

*Stigler, George J. (1951): The Division of Labor is Determined by the Extent of the Market. In Journal of Political Economy, 59 (3) pp. 185-193*

*Thomke, Stefan and Hippel, Eric von (2002): Customers as Innovators: A New Way to Create Value. In Harvard Business Review, 80 (4) pp. 74-81*

*Urban, Glen L and Hippel, Eric von (1988): Lead User Analyses for the Development of New Industrial Products. In Management Science, 34 (5) pp. 569-582*

*Varian, Hal R. (2002). New Chips Can Keep a Tight Rein on Consumers. Retrieved [Date unavailable] from NYTimes: http://www.nytimes.com/2002/07/04/business/04SCEN.html?pagewanted=all*

## About the author

### Eric von Hippel

*© Eric von Hippel*

Has also published under the name of:
"Eric von. Hippel"

I am a Professor of Technological Innovation in the MIT Sloan School of Management, and am also a Professor in MIT's Engineering Systems Division. I specialize in research related to the nature and economics of distributed and open innovation.

I also develop and teach about practical methods that individuals, open user communities, and firms can apply to improve their product and service development processes.

## Your notes and thoughts on chapter 18

Record your notes and thoughts on this chapter. If you want to share these thoughts with others online, go to the bottom of the page at:
http://www.interaction-design.org/encyclopedia/open_user_innovation.html

**NOTES**

# Chapter 19

# Visual Aesthetics

## in human-computer interaction and interaction design

*by Noam Tractinsky.*

Visual aesthetics, as discussed in this chapter, refers to the beauty or the pleasing appearance of things. We discuss the importance of visual aesthetics in the context of interactive systems and products, present how it has been studied in the field of Human-Computer Interaction (HCI), and suggest directions for future work in this field.

## 19.1 INTRODUCTION

To scholars and practitioners in the field of HCI at the early 1990's, the idea that aesthetics matter in information technology sounded heretic. Two decades later, in the early 2010s, this thought has conquered a solid place in both academia and industry. While experimentation with computers' ability to generate visual art dates back to the 1960's (Nake, 2005), systematic research on visual aesthetics of interactive systems can only be traced to the mid-1990's (Kurosu and Kashimura, 1995; Tractinsky, 1997). Since then, a steady stream of studies has explored various aspects of this area. The timeline of this research has roughly corresponded to even more dramatic developments in the information technology industry. Since the later 1990's, a strong shift towards visual aesthetics has swarmed the industry. The increased interest in aesthetics among the industrial and academic communities reflects the maturation of the HCI field and the overcoming of many of its growing pains as a discipline that struggles with unreliable technology on the one hand and with the need to satisfy users' basic requirements on the other hand. Additionally, broader societal processes emphasizing design and style emerged at about the same time (Gibney and Luscombe, 2000; Postrel, 2002), further reinforcing shifts towards aesthetics of products in general (Bloch, 2011) and specifically of interactive systems. A more detailed account of this process is provided in Tractinsky (2004) and Tractinsky (2006).

Udsen and Jørgensen (2005) identified several approaches to the study of aesthetics in HCI. "Visual aesthetics", as described in this chapter, correspond roughly to the approach which Udsen and Jørgensen identified as "Functionalist". To be specific, and to distinguish the subject of this chapter from other similar terms, I use the term "aesthetics" in its fairly ordinary and common sense as reflected in dictionary definitions such as "an artistically beautiful or pleasing appearance" (The American Heritage Dictionary of the English Language), or as "a pleasing appearance or effect: Beauty" (Merriam-Webster's Collegiate Dictionary). The term "visual" indicates concentration on the visual sense, which is the central human sense, occupying "almost half the brain" (Ware, 2008, ix). Thus, this chapter is *not* about various other phenomena studied under the "aesthetics" heading, such as literary aesthetics, abstract forms of aesthetic experiences or criteria (e.g., the elegance of mathematical proofs), or reactions to object qualities that do not immediately and primarily stem from its visual attributes.

In addition, a few other characteristics that describe research in the field can be listed. These characteristics describe how researchers in the field approach their subject matter. First, the approach of researchers in visual aesthetics reveals a bias towards positive effects of visual design, an issue to which I will return later in this chapter. Hence, research in this area commonly studies the beautiful and pleasing appearance of artifacts, or designed objects that are based on computing technology, rather than the effects of their ugly and displeasing counterparts. Second, at a Dagstuhl workshop on visual aesthetics in HCI, held in 2008, a majority of the participants adopted an interactionist approach to the study of visual aesthetics, noting that the aesthetic experience consists of people's reactions to objects as opposed to aesthetics that are inherent in the object per se (Hassenzahl et al., 2008). These reactions include both individual idiosyncrasies and tastes, but also considerable agreement between individuals and experts, to a point where they may be considered "quasi-objective" (Hoyer and Stokburger-Sauer, 2011). Third, while the Dagstuhl workshop mentioned above failed to reach a consensus over the time frame that appears relevant to visual aesthetic reactions, my own position is that it can encompass the entire range from very quick, visceral reactions to very long, contemplative evaluations. Fourth, the processes involved in designing and evaluating visual aesthetics are both affective and cognitive. Finally, research in the field of visual aesthetics is primarily empirical and is characteristically descriptive (i.e., "what *is* considered beautiful") rather than normative (i.e., what *should* be considered "beautiful") (Hassenzahl, 2004b). This important distinction stresses its roots in applied research and differentiates the field from artistic or philosophical writing on the subject.

The objective of this chapter is to survey the field of visual aesthetics in HCI. We start by delineating the importance of visual aesthetics to HCI from three perspectives. We then present a research framework that serves us in reviewing key findings in the field. These aspects include issues such as what makes systems look aesthetic, what are the effects of visually aesthetic systems, and what mechanisms are involved in people's judgment of aesthetics in the context of interactive systems. We also discuss methodological aspects and challenges for further research.

### 19.1.1 The importance of visual aesthetics in HCI

The importance of visual aesthetics to the field of HCI can be argued from various perspectives. Here I present three such perspectives – the design perspective, the psychological perspective, and the practical perspective. Although these perspectives are not meant to be exhaustive, I believe that, taken together they cover the lion share of arguments for the inclusion of visual aesthetics as a major aspect of HCI practice, research and education (Tractinsky and Hassenzahl, 2005). While these are distinctive perspectives, they may overlap at certain points. Finally, to some it may seem somewhat redundant to submit these arguments, as they have gained considerable acceptance in the HCI community in recent years. Still, I believe that it is important to present them in an organized fashion to clarify my point of view and to provide people in the community a set of arguments that can be used to make a case for visual aesthetics before other audiences (e.g., software and hardware engineers, product managers, etc.).

### 19.1.2 The Design Perspective

Despite being one of the youngest design disciplines, the development of interactive technology has quickly become one of the most salient design activities. The mutual relevance of HCI and Design has long been recognized (e.g., Winograd, 1996), but what are the implications of this to the research, practice and education in HCI? Here, I would like to point out two such implications. The first implication is the recognition that aesthetics constitutes an important and integral part of any design discipline. The importance of aesthetics increases as the interface between the artifact and the affected people (e.g., in terms of visual saliency, length of interaction or co-habitation) becomes more comprehensive. The second implication is that visual aesthetics is often related to other design aspects. Thus, not only should we not worry about trading off aesthetic and other qualities of interactive systems; we should embrace aesthetics as a dimension that augments other aspects of the design and the overall interactive experience.

#### 19.1.2.1 The Vitruvian design principles

Vitruvius (1st century BC) is probably the first person to lay forth systematic and elaborated principles of design. It is not surprising that architecture was the subject of his elaborated writings, being the most salient and complex design discipline, which has affected human life ubiquitously. In addition to the fact that information technology and interactive systems have now become just as ubiquitous, it is not difficult to see that there is much in common for architecture and information technology (e.g., Brooks, 1975; Hooper, 1986; Lee, 1991; Kim et al., 2002; Visser, 2009). It is reflected by the term "information architecture," used by professionals to designate the process of creating information-based environments and systems. The similarities between these two disciplines can be illustrated by considering Vitruvius's three core principles of sound architectural work. *Firmitas*, which is the strength and durability of the building; *utilitas* – the utility of the building, its usefulness and its suitability for the needs of its intended inhabitants and users; and *venustas* – the building's beauty. In architecture, the Vitruvian principles have been influential since their rediscovery in the 15th century (Johnson, 1994; Kruft, 1994). Today, for example, they serve as a basis for the Design Quality Indicator, developed by the Construction Industry Council in the U.K. to evaluate the design quality of buildings (Whyte et al, 2003).

It is straightforward for the various computing and IT disciplines to recognize *firmitas* as the core principle of their research and practice. The need for robust, reliable and dependable software, hardware, systems and products has occupied the field since its inception. We might say that, just as *firmitas* serves as a prerequisite for designing structure, so do we consider it a precondition for any IT system or product.

Whereas there is little disagreement about the importance of *firmitas* principle, the computing community was originally much less enthused about the *utilitas* principle. In the context of IT, this principle deals with designing to meet individual and organizational needs and goals, with emphasis on the efficiency and the effectiveness of the interaction between people and artifacts. In fact, the HCI community can take much of the credit for incorporating the *utilitas* principle into mainstream practices in the computing industry (cf. Tractinsky, 2006). The field of HCI has its roots in attempts to study and design systems and product that will allow people to use them efficiently (Card et al, 1983). The notion of usability, for example, which has served as a centerpiece of the HCI community has permeated not only other parts of the IT industry, but have gained almost universal recognition and support for the values of human-centered design. One of the most widely referenced models on people's relationships with information technology – the technology acceptance model (TAM) – suggests that a system which is easy to use and provides more useful features is more likely to be adopted by its intended users (Davis, 1989).

With *firmitas* and *utilitas* in place, the computing community in general, and the area of HCI in particular, are still missing a key Vitruvian principle. For years, beauty and delight were considered by the HCI community as gratuity, often to be avoided. The emergence of beautiful interactive products during the first decade of the 21st century , which led to commercial success and to academic research (e.g., Kim et al., 2002; Liu, 2003; Tractinsky, 2004), has demonstrated quite convincingly, that as in other design disciplines, the third Vitruvian leg, *venustas*, should be fully embraced as cornerstone of designing interactive technology (see also, Fishwick, 2006).

Figure 19.1: Italian translation from 1521 of De Architectura Libri Decem (The Ten Books on Architecture) by Marcus Vitruvius Pollio. Preserved in the Smithsonian Museum of American History.

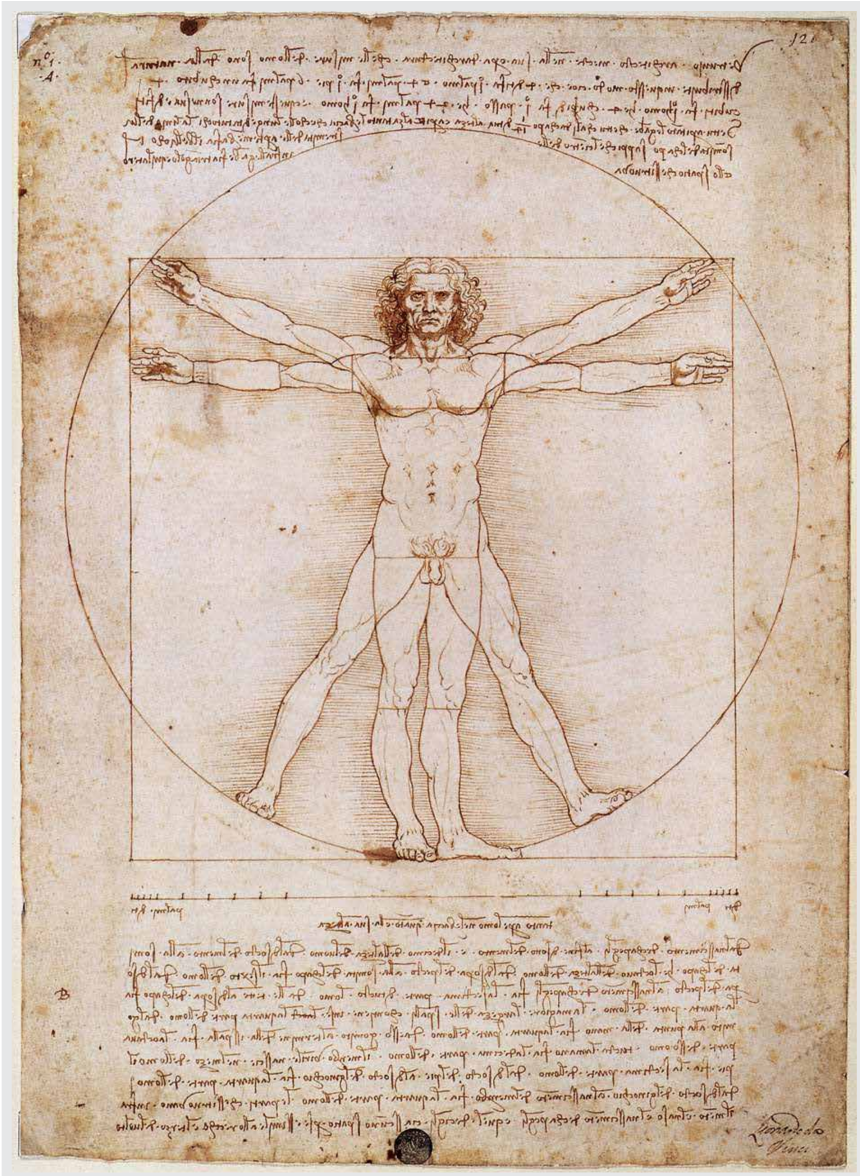


Figure 19.2: The Vitruvian Man drawing was created by Leonardo da Vinci circa 1487 based on the work of Vitruvius. By empirically measuring and calculating the proportions of the human body, Vitruvius may also be considered the first student of ergonomics.

#### 19.1.2.2 Aesthetics and other design principles overlap

HCI and usability experts used to warn against putting too much emphasis on aesthetics (e.g., Norman, 1988; Nielsen, 1993). Their warnings seem to reflect a concern about the ability of these two design aspects to coexist. Beauty was a hurdle on the road to good design. If designers emphasize aesthetics they by default sacrifice usability. This viewpoint has been changing gradually, thanks in part to a stream of research findings that suggest that at least in terms of perceived design attributes, aesthetics and usability can be viewed as positively correlated (Tractinsky et al., 2000; Lavie and Tractinsky, 2004, Cawthon and Moere, 2007; Sonderegger and Sauer, 2010). Moreover, a closer look at usability guidelines suggests that there is no inherent conflict between usability and aesthetic principles. Guidelines for usable computer applications rely heavily on Gestalt laws of perception in recommending, for example, orderly displays, keeping elements aligned, grouping elements that belong together, clearly separating them from other elements, etc.

Of course, these principles were applied as well to explain and promote the theory and practice of art and design, suggesting that they affect aesthetic impressions (Arnheim, 1966). One of my favorite demonstrations of this point is the following screens which appeared in a study by Parush et al (1998). Participants in that study were asked to evaluate an interface quality of two screens (Figure 19.3). The participants rated the quality of the left screen as better than the quality of the right screen. But which design quality were they referring to? Usability? Visual aesthetics? Time and again, when I pose this question during class or in invited lectures, the distribution of answers remain almost the same: They are evenly distributed between "usability," "aesthetics" and "both"!

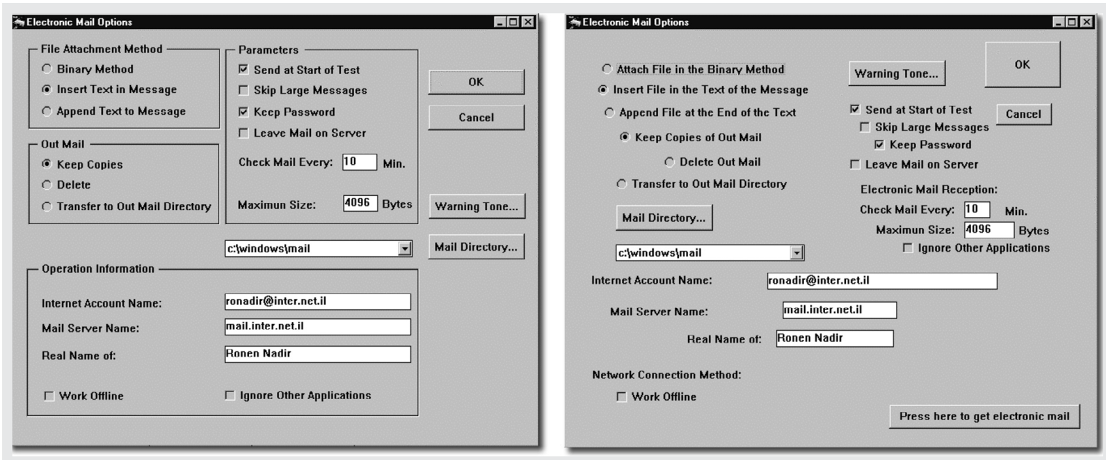



FIGURE 19.3: Two screens from Parush et al (1998). The left screen represents good design. The right screen represents bad design.



This example illustrates the findings of another study (Lavie and Tractinsky, 2004) in which web pages were characterized along two perceived dimensions. We used the terms "classical" aesthetics to denote the dimension that communicated a sense of order and good proportions. This subdimension was highly correlated with usability. The other dimension represented originality and creativity aspects of the design and was accordingly labeled "expressive" aesthetics. Not surprisingly, perceived website usability was highly correlated with classical aesthetics but only moderately with expressive aesthetics. Thus, it is important to realize that (a) people can distinguish among different aesthetic aspects of interactive systems, and that (b) at least some aspects enhance, rather than negate, usability.

### 19.1.3 The Psychological Perspective

In the early days of the HCI discipline, researchers and practitioners emphasized task-related criteria, such as ease of use and efficiency as the ultimate goals of interactive design. Aesthetics was considered gratuitous at best or even harmful (e.g., Norman, 1988). However, as interactive technology became so ingrained in everyday life, touching on almost all aspects of it, it became apparent that this position should be reevaluated (Norman, 2002, Hassenzahl, 2007). To a large extent, the emergence of visual aesthetic research in HCI had its roots in the “positive psychology” movement (Seligman and Csikszentmihalyi, 2000) that called for a shift towards dealing with human strengths and well-being instead of with weaknesses and their remedies. This sentiment was enthusiastically embraced in the field of HCI in the context of studying the user experience (Hassenzahl and Tractinsky, 2006; Law and Schaik, 2010)

This section provides three arguments, from a psychological perspective, for the importance of aesthetic design. The basic idea is that aesthetic design positively influences both emotional and cognitive processes (Norman, 2004; Leder et al., 2004). This, in turn improves people’s experience with the technology, their appraisal of it and their attitudes towards it (e.g., Hartmann et al., 2007, 2008; Thuring and Mahlke, 2007). In this chapter we first discuss the emotional and motivational aspects: aesthetics pleases us and improve our well-being. We then discuss cognitive processes by which visual stimuli are easily recognized and thus are essential to subsequent evaluation of products and environments.

#### 19.1.3.1 Aesthetics satisfies basic human needs and is a source of pleasure

Early HCI writings appear to belittle the need for aesthetic design. Such a perspective may have been motivated mainly by the need to promote the more pressing values of usable design. Still, given our knowledge about human nature, this position was not sustainable in the long run. It is argued that the value of visual aesthetics stems from its contribution to pleasure and well-being (e.g., Santayana, 1896; Postrel, 2002), and from its role as a basic human need (Maslow, 1954), perhaps due to evolutionary processes (Dutton, 2008).

FIGURE 19.4: Aesthetics as an extension of the Self: Harry Potter skin for a Blackberry smartphone.



Visual aesthetics may temporarily take a side seat to other design aspects when other needs are more pressing; some people may be less sensitive or less in a need for aesthetic environments (Bloch et al., 2003); and aesthetic tastes, reactions to aesthetic stimuli vary between people (Santayana, 1896; Hoyer and Stokburger-Sauer, 2011). Still the universality of visual arts across cultures and the pleasures induced by it are cited by evolutionary psychologists as evidence for the fundamental role of aesthetics in the psyche of modern *Homo sapience* (Dutton, 2008). Aesthetic experiences are associated with affective responses and reflective thought (Leder et al., 2004). Research using functional magnetic resonance imaging (fMRI) found further neurophysiologic support for this association in the context of product packaging (Reimann et al., 2010). Whereas task-related criteria are often based on extrinsic motivation, aesthetics, through pleasure and engagement, primarily contributes to intrinsic motivation. Thus, there is little reason to believe that the need for aesthetics disappears in front of the computer. Visually pleasing design enriches our experiences with interactive systems just like they do with any other environment (Hassenzahl, 2007). There is empirical evidence that aesthetic design of interactive technology increases users' pleasure and engagement (e.g., Thuring and Mahlke, 2007; Porat and Tractinsky, 2012; Angeli et al., 2006). Consequently, we expect pleasurable interactions to make us happier and thus to improve our well-being (Lyubomirsky et al., 2005). Furthermore, they may make us more tolerable of other design imperfections (Norman, 2004) and improve our task performance under certain conditions (Moshagen et al., 2009).

#### 19.1.3.2 Aesthetics as an extension of the Self

Belk (1988) argues that "it seems an inescapable fact of modern life that we learn, define, and remind ourselves who we are by our possessions" (p. 160) This process of self-extension by possessions is manifested in the realm of HCI in the ways that people modify their computer desktops, screen savers, their websites, and various standard applications. These forms of idiosyncratic attachments (Kleine et al., 1995) are served well by the flexibility and plasticity of IT. The appearance of modern software applications can be personalized to suit users' tastes. This trend, though on a smaller scale, can be found in hardware – e.g., in the abundance of various cell phone covers, charms and other ornaments. Software skins can be downloaded, for free or pay, for most popular applications. Indeed, studies have shown that the major factor influencing users' selection of skins was the aesthetic aspect of its design (Tractinsky and Lavie, 2002; Tractinsky and Zmiri, 2006). The proliferation of such features could be explained in large part by individuals' desire to express themselves, and to be seen in specific ways by others (Hassenzahl, 2003) and as part of an ongoing process of identity formation, through which people "express who they are and communicate affiliation with others" (Venkatesh and Meamber, 2008, p. 51). These are manifestations not only of who those individuals are, their past and present, and their affiliations (Kleine et al., 1995) but also of the social context within which so many interactive products are used today (Turkle, 2005).


FIGURE 19.5: Aesthetics as an extension of the Self: Harry Potter skin for the Windows 7 operating system.


Figure 19.6: Aesthetics as an extension of the Self: Harry Potter skin for the Google Chrome browser.



### 19.1.3.3 Aesthetic impressions are fast, enduring and consequential

Whereas the previous arguments discuss psychological needs for aesthetic environments and motivation to possess, buy and use aesthetic products, the current argument is based on the consequences of aesthetic stimuli. Those consequences are based on the idea that aesthetic impressions can be very fast. Studies of brain activity suggests that aesthetic impressions form within 300ms to 600ms (Höfel and Jacobsen, 2007). Research on people's impressions of web pages demonstrated that reliable and consistent aesthetic judgments are formed with exposure of less than 500 milliseconds (Lindgaard et al., 2006; Tractinsky et al., 2006).

These very fast impressions are the first opportunity we have to form an attitude towards an object (e.g., an interactive system), whose other qualities are usually concealed until later time when opportunities to evaluate them arise (e.g., when trying to accomplish a task with the system). Those initial attitudes are likely to form at a relatively subconscious level and therefore may be relatively uniform across people, relative to more elaborated evaluations (Kumara and Gargb, 2010).

The primacy of first impression on attitudes is well documented in social science research. Its most salient manifestation is expressed by the "what is beautiful is good" stereotype (Dion et al., 1972), which suggests that a person's physical appearance affects how others view the person's hidden qualities (e.g., personality traits). Such preferences for aesthetic appearance may be the result of evolutionary adaptation (Rhodes, 2006). Research has documented numerous contexts in which people with good looks enjoy preferential treatment in the labor market (Hamermesh and Biddle, 1994), in credit markets (Ravina, 2008), and even in the classroom (Hamermesh and Parker, 2003).

We also know that people try to actively improve how they appear to others in order to gain benefits or to avoid sanctions (Jones, 1990). Such attempts can be found, for example, in how people try to improve information about things under their responsibility (e.g., at work) by presenting the information in more attractive formats (Tractinsky and Meyer, 1999). Such aesthetic improvements may pay off: research suggests that under ordinary conditions, aesthetic financial reports increase both novice and professional investors' valuation of a firm (Townsend and Shu, 2010). Similarly, the way things appear may influence our attitudes towards them. By "things" we may refer to natural settings and objects such as landscapes (Porteous, 1996; Carlson, 2000) or to various sorts of designed environments (Gilboa and Rafaeli, 2003) and artifacts (Rafaeli and Vilnai-Yavetz, 2004). For example, Reimann et al (2010) found that products with aesthetic packages are chosen over less expensive products with well-known brands in standardized packages.

Thus, it should come as no surprise that the visual aesthetics of interactive systems, both hardware and software, may affect our evaluation of other system attributes. Hence, the suggestion that “beautiful is usable” that is, beautiful systems are considered by users to be more usable (Tractinsky et al., 2000).

### 19.1.4 The Practical Perspective

Finally, I would like to suggest that even if you are not convinced that aesthetics is a pillar of good design or that aesthetics fulfill psychological needs and influence attitudes and decision making, its practical significance is hard to deny. There is ample daily life evidence that illustrate this point. Here, I would like to suggest two aspects of this perspective. The first describes the importance of aesthetics as a differentiating factor between similar products. The second argument suggests that aesthetics and information technology are already profoundly intertwined in current socio-technical processes. Thus, not only is it unwise to ignore aesthetics in information technology, we are in fact required to pay more attention to it, to further study its relationships with other relevant aspects of the HCI field, and to improve its integration in the design of interactive systems.

#### 19.1.4.1 Aesthetics as a differentiating factor

The weight of the IT industry has shifted over the last fifty years from emphasizing organizational number crunching to supporting organizational and personal decision making and productivity to being fully integrated in consumer and entertainment products. The list of successive companies that dominated the IT industry during this time frame – IBM, Microsoft and Apple – tells the story succinctly. The accelerated process of consumer-centeredness and commoditization of interactive technologies, partially described already byNorman (1998), increases the importance of aesthetics as a differentiating factor between competing products increases. The digital watch industry served as a classic example of such a process during the 1970's and 1980's, as functionality and performance met very high standards of accuracy and reliability. In that industry, much of the differentiation between brands and models is now based on aesthetic creativity or imitations of aesthetic exemplars. Today we are surrounded by similarly high-performance interactive consumer products – such as smart phones and tablet computers. These products are more oriented towards enhancing the user experience, and much of the battle involves attempts to catch the consumer's eye and heart with appearance and design-based symbolic value. Thus, aesthetic design is gaining acceptance as a differentiating strategy or tactic (Simonson and Schmitt, 1997; Luchs and Swan, 2011; Reimann et al., 2011) in various markets.

Figure 19.7: The first version of the iPhone, released in 2007 (left) compared to its contemporaries. The iPhone is a good example of how a phone manufacturer uses visual aesthetics as a differentiating factor - in everything from the actual phone to its packaging.

*Courtesy of Marco Arment. Copyright: CC-Att-SA-3 (Creative Commons Attribution-ShareAlike 3.0).*

FIGURE 19.8.A: The hugely popular Western Electric Model 2500 (12 button Touch-Tone) telephone, manufactured in 1980.



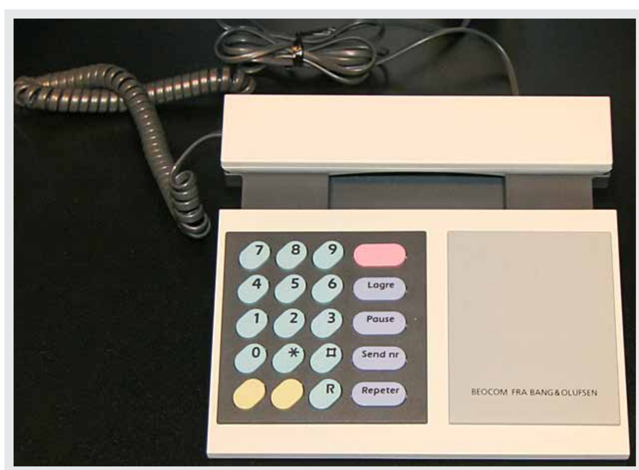



FIGURE 19.8.B: The BeoCom 1000 corded analogue telephone used visual aesthetics to differentiate itself from and compete against popular telephones like the Western Electric Model 2500.

### 19.1.4.2 Aesthetics is pervasive

There is a long tradition of relationship between aesthetics and technology (Petroski, 1992; Norman, 2004), which became more pronounced during the rapid technological developments of the 20th century. Advanced technologies helped in producing aesthetic forms that could not have been made before (at least not on a mass scale), and aesthetic concepts were borrowed from one technological field (e.g., aircraft design) into another (e.g., locomotive design or a structure of an airport terminal) with the aid of new design, manufacturing or building technologies. Today, it is common for design trends to appear at the same time in multiple industries (Gladwell, 2000). In the age of information technology, such relationships become more symbiotic than ever. One of the unique characteristics of information technology is that it is particularly friendly to aesthetic applications (Postrel, 2002). Relative to traditional methods it supports effortless creation, manipulation, and transmission of aesthetic materials. Consider, for example, the photography market's shift from film-based cameras to digital cameras within merely a decade. Today, images and photos, major elements of visual aesthetics comprise about 2/3 of the volume of web pages transmitted through the internet (Rabbat, 2010). Digital technology and applications enable designers in various industries many more design options, and much more time to explore all of those options in order to create more appealing products. Perhaps even more importantly, these technologies offer ordinary people tools that help creating and communicating aesthetics on a revolutionary scale. The interlacing of aesthetics and information technology thus creates an aesthetic cycle in which constant supply of visually aesthetic stimuli increases people's aesthetic sensitivity, which in turn motivates people to seek aesthetics everywhere, including in interactive products (Postrel, 2002).

### 19.1.4.3 A note on the moral aspect of practical considerations

Lest our position be misconstrued to mean that advocating aesthetic design implies that practical concerns should prevail over all other considerations, it is important to note that the reality depicted above also carries a moral dimension. Aesthetic and ethic considerations in design need not contradict (Liu, 2003). Aesthetic design implies that the designer or the organization respects their audience, is sensitive to people's needs and desires and puts thought and effort into the design of the product and the environment. For example, Ulrich's (Ulrich 1984) seminal study on hospital patients recovering after cholecystectomy found that patients in rooms with window view of natural settings recovered faster and needed fewer potent analgesics relative to patients in rooms with windows facing a brick building wall. A quarter of a century later, Postrel (2008) laments the disregards for aesthetic design in today's health care institutions. In turn, feeling respected and appreciated, people will be more inclined to take care of an aesthetic and well maintained environment (Saito, 2008). In short, aesthetic design works for the betterment of our lives.

## 19.2 RESEARCH ON VISUAL AESTHETICS IN HCI

Research on and around the various aspects of visual aesthetics in HCI can be roughly divided into four main categories that deal with various aspects of the aesthetic process:

1. Antecedents of the aesthetic evaluation, that is, what makes people engage in aesthetic evaluations, and perhaps more importantly, what causes variations in aesthetic evaluations;
2. The aesthetic evaluation itself and the psychological processes that are involved in it;
3. Outcomes or consequences of aesthetic evaluations;
4. Moderating variables, or intervening factors that influence how the antecedence operate on the aesthetic evaluations and how the evaluations affect the outcomes.

These categories can be seen as part of the schematic process described in Figure 19.8. I will elaborate on each of these elements below and present empirical studies that have handled these issues.

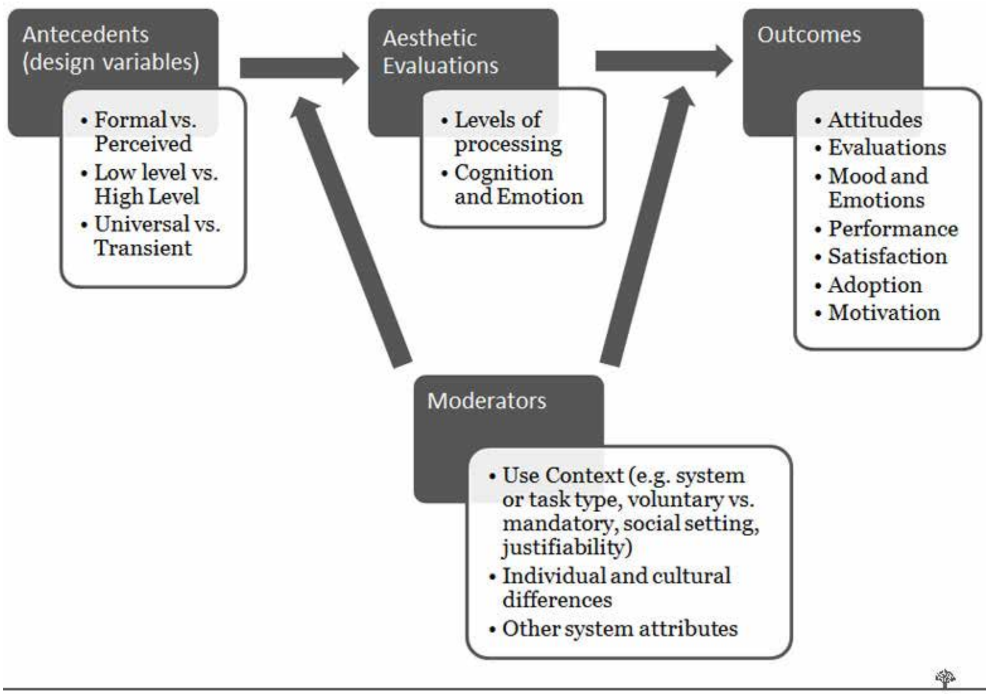



Figure 19.9: A general framework for the study of visual aesthetics in HCI. Some relevant research issues for each category are included.



### 19.2.1 Antecedents of visual aesthetics

What makes an interactive system look more or less aesthetic? What makes one system look aesthetic in a particular way and another system look aesthetic in a different way? These are questions of the utmost practical importance. If we could only decipher the aesthetic code! Fortunately, the quest for the Holy Grail of visual aesthetics is beyond our reach for at least the foreseeable future, so there is plenty of room for experimentation, new approaches and ample research. In studying this category, we naturally look first at design guidelines and insights. However, the very broad scope of design possibilities, the creative nature of design work, and the almost unbounded relationships between design elements make it extremely difficult to isolate specific design aspects which may be considered aesthetic or which may influence aesthetic perceptions one way or another. It is probably possible to categorize aesthetic design guidelines from the very broad and abstract (e.g., "form follows function") to the very specific (e.g., "use colours with different hues between background and menu bar" (Kim et al., 2003), with mid-range guidelines in between (e.g., "visual layout should be symmetrical" (Sutcliffe, 2002). They can be expressed in terms of object properties or in terms of motivational and emotional mediators (e.g., Berlyne's (Berlyne 1971) collative variables).

At the beginning of this section, we posed two different questions. The first question -- what makes a system look more or less aesthetic -- has probably been more central to aesthetic and design research over the years. Park et al. (2005) have collected and listed 11 visual attributes that can potentially answer this question. Other, more high-level responses to this question include contrasting attributes such as novelty and typicality (Veryzer and Hutchinson, 1998; Hekkert et al., 2010) and the related idea of processing fluency (Reber et al., 2004). Hekkert et al.'s results suggest that a balanced dose of typicality and novelty increase aesthetic evaluations (see also Kumara and Gargb, 2010, Tractinsky et al., 2011a). Similarly, van Schaik and Ling (2011) suggest that design qualities, which they term pragmatic and hedonic, affect perceptions of overall beauty.

Some researchers argue for the prospect of identifying formal, objective, attributes that determine aesthetic judgment, and which will ultimately lead to automatic composition or checks of displays such as web pages (e.g., Ngo et al., 2003). This approach has been criticized on the grounds that aesthetic laws engrained in the object are "universalist" (Krippendorff, 2005) would not survive individual, cultural and context differences (Martindale et al., 1990; Krippendorff, 2005). Similarly, Csikszentmihalyi (1991) argues that formal aspects only rarely make objects valuable to their owners. He speculates that people do not perceive formal attributes such as order or disorder in design according to mathematical principles. Still, despite the apparent subjective and context-dependent nature of aesthetic processes, studies have continued the quest for basic and formal principles of aesthetic properties of interactive systems. Such principles can be expressed as computational models aimed at achieving optimal design spaces. For example, Bauerly and Liu (2006) suggest that in basic images, symmetry and balance affect aesthetic appeal ratings. However, they also found that the strong relationship found between symmetry and aesthetic appeal diminished when tested with more realistic (i.e., context-dependent) web pages. Other approaches to predicting aesthetic evaluations have been proposed in the context of photographs. For example, the Aquine project (Datta et al., 2006, Datta et al., 2008) proposes to combine various algorithmic approaches to classifying photographs according to various visual properties.

However, as mentioned above, the problem of finding universal visual aesthetic guidelines and laws is further exacerbated in the field of HCI because of the variety of applications and products and the uniqueness of so many use contexts. In addition, the dynamic nature of contemporary society and fashion-like approach to the design of many interactive devices and applications make aesthetics a moving and often unpredictable target (Korman-Golander, 2011).

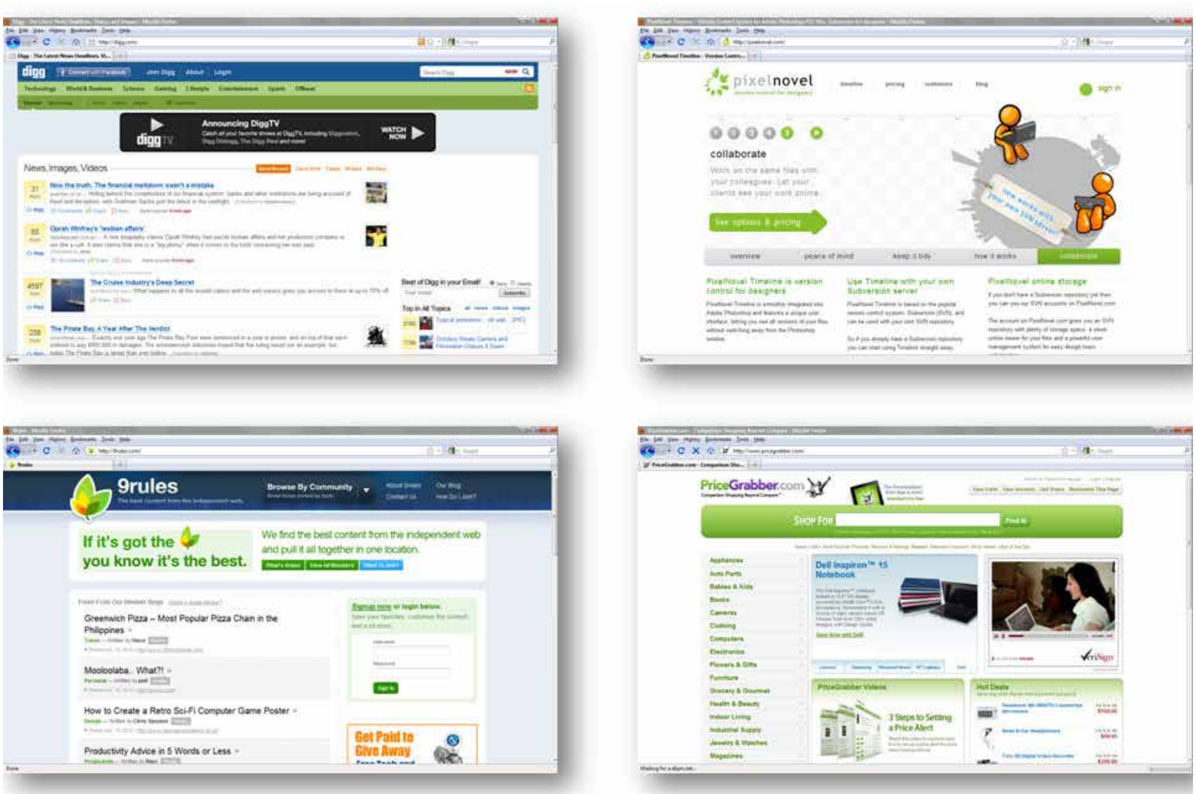



Figure 19.10: Website design fashion changes continuously. The popularity of the "Web 2.0" design trend peaked around 2007 and has been on the decline since then (Korman-Golander, 2011).

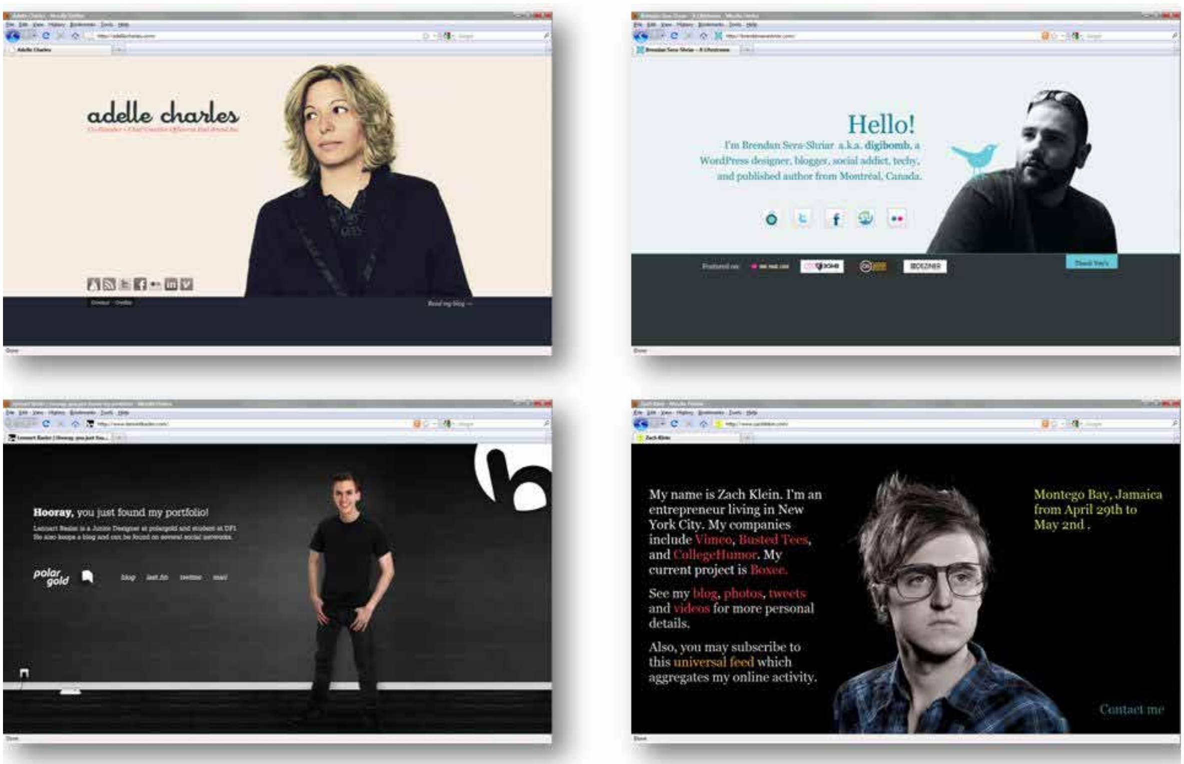



FIGURE 19.11: The "One-Page Layout" website design trend became popular in 2008 (Korman-Golander, 2011).



These constraints lead us to the second question. Here, researchers have tried to break down the aesthetic stimuli to sub-dimensions which may be more or less suitable for various contexts or to individual tastes. Such dimensions often emerged out of subjective evaluations of aesthetic stimuli. For example, Park et al. (2004) identified thirteen aesthetic dimensions of web pages. A more parsimonious approach to dimensionality was taken by Lavie and Tractinsky (2004), who identified two perceived dimensions of visual aesthetic, "classical" and "expressive," in the context of website design. Classical aesthetics corresponds to traditional views of aesthetic – symmetrical, clean and organized design. The expressive dimension relates to the designer's creativity and originality. One of the important aspects of that study was the demonstration that these aesthetic dimensions are correlated as expected with various interaction outcomes such as perceived usability, pleasing interaction and perception of service quality. Similarly, Moshagen and Thielsch (2010) have suggested four aesthetic dimensions: Simplicity, diversity, colorfulness and craftsmanship. The first two dimensions were highly correlated with Lavie and Tractinsky's classical and expressive aesthetics, respectively. All four dimensions were associated with appeal. Commensurate with Lavie and Tractinsky's results, all dimensions were positively, but differentially correlated with various outcome measures.

### 19.2.2 Perceiving and evaluating visual aesthetics

This category deals with one of the most basic questions in the field of visual aesthetics: How people process and evaluate visual stimuli in aesthetic terms? Detailed accounts of such processes are likely not specific to HCI, and thus have been and probably will be left to researchers in more basic research fields. Findings from such research are presented below to inform the readers about developments in this field. One of the most influential accounts of aesthetic processes was articulated by Norman (2004), who suggested that aesthetic perceptions and evaluations can be explained by considering cognitive and emotional processes at three different levels, which he termed visceral, behavioral and reflective. Visceral reactions to stimuli in the environment (including aesthetic stimuli) have developed to a large extent through evolutionary mechanisms, are performed very rapidly at almost instinct level, with little or no cognitive processing (Ortony et al., 2005). Thus, reactions at this level are quite automatic. The other two levels are characterized by increasingly more elaborated and distinct motivational, emotional and cognitive structures and processes, as well as by slower reactions to stimuli, tendency towards more optimal (as opposed to satisficing) responses and greater individual variability (Ortony et al., 2005).

Studies of aesthetic reaction have been performed on all levels (e.g., Csikszentmihalyi 1991; Leder et al., 2004; Winkielman et al., 2006; Jacobsen, 2010). Low level research is characterized by processes that last a few tenths of a second. At this level, research suggests that the evaluative aesthetic judgment involves a two-step process of an early impression formation and a later evaluative categorization process (Höfel and Jacobsen, 2007). Another finding at this level argues for a positive effect of prototypicality on aesthetic evaluations through the ease (fluency) of information processing (Winkielman et al., 2006). In the field of HCI, several studies have examined aesthetic evaluations after very short exposure to web pages. One of the differences between the basic research and HCI research at this level is that the latter usually involves more ecologically representative (i.e., "real") stimuli. Lindgaard and colleagues (Lindgaard et al., 2011) suggest that stable aesthetic evaluations can be formed even after being exposed to a design for only 50 milliseconds. While some research questions the robustness of these findings, other research supports the notion that we do not need more than half a second to form first, and stable, aesthetic impressions of the web page (Lindgaard et al., 2006; Tractinsky et al., 2006)

Research on aesthetic processing at higher levels involves more elaborated considerations. In general, Leder et al. (2004) have proposed a model of aesthetic appreciation and aesthetic judgment. The model includes various categories of aesthetic processing, including "automatic" and "deliberate" stages and cognitive and emotional reactions. Although studying higher-level processes of aesthetic evaluations may be of interest to the HCI community, research thus far has concentrated more on the role of aesthetic processing as a mediator between design stimuli and outcome variables such as user engagement and trust (e.g., Hartmann et al., 2008, Lindgaard et al., 2011). Similarly, Thuring and Mahlke (2007) propose a model which integrates the effects of perceived system qualities, including visual aesthetics, on emotions and on appraisal of the system. Research on more long-term, reflective aesthetic evaluation is even scarcer. An example for such research in a general context is Csikszentmihalyi's (Csikszentmihalyi 1991) study on household objects. Recently studies on aesthetic aspects of interactive systems began to adopt a more time-dependent approach (Schaik and Ling, 2008) and to employ longitudinal methods (e.g., Karapanos et al., 2010).

### 19.2.3 Outcome Variables

This section deals with what is probably the most practical question in relation to visual aesthetics in HCI: What are the effects of aesthetic perceptions and evaluations on HCI-related variables? Whereas the question of how aesthetic evaluations are formed is relatively general, the consequences of these evaluations can be highly domain-specific. Indeed, research in HCI primarily views the value of visual aesthetics, whether explicitly or implicitly, not as an end in itself but rather as a mediating force between (1) characteristics of the designed system or product, and (2a) other perceived attributes of the product or (2b) behavioral consequences of aesthetic evaluations. Early studies (Kurasu and Kashimura, 1995, Tractinsky, 1997) offered intriguing findings regarding positive association between visual aesthetics and a focal HCI variable - usability. Despite the lack of clear indication about cause and effect, these studies hinted at what subsequent studies explored more directly: that people's perceptions of a system's beauty may impact perceptions of other system attributes, such as ease of use (Heijden, 2003; Cyr et al., 2006), overall satisfaction (Tractinsky et al., 2000; Lindgaard and Dudek, 2003; Cyr, 2008), preferences (Schmidt and Liu, 2009; Lee and Koubek, 2010) and even performance (Quinn and Tran, 2010, Sonderegger and Sauer, 2010).

While early research appeared to accept intuitively the premise that aesthetic evaluations are fast enough to precede evaluations of other related variables, later research has demonstrated that this is indeed the case (Lindgaard et al., 2006, Lindgaard et al. 2011; Tractinsky et al., 2006; see also Section 2.2 above). These findings have further motivated the exploration of potential consequences of aesthetic design, a research area which have probably been the busiest of the four categories outlined in our framework. The range of variables covered by these studies includes evaluations of other system attributes, overall evaluations of the system, attitudes towards organizations represented by the system, and satisfaction from the interaction. Studies have also begun exploring the role of affect and emotions in mediating how perceived aesthetics influence those outcomes. Below I survey several studies that examined visual aesthetics' effects on various variables. The studies represent a partial and probably arbitrary list.

*Affect* and *emotions* are oft-cited corollaries of visual aesthetics. The effects of attractive and appealing design on emotions were demonstrated in various studies, including Thuring and Mahlke (2007) in the domain of portable music players, Porat and Tractinsky (2012) and Cai and Xu (2011) in the domain of online shopping. The importance of aesthetics' effects on emotions is twofold. First, as mentioned earlier, positive affect contributes to positive experience and well-being, and as such is an end in itself. Second, emotions have a role in affecting subsequent information processing, appraisal of other system attributes, and forming attitudes towards the system (Sun and Zhang, 2006).

*Trustworthiness* was a variable that was studied early as an outcome of visual design in the domain of online banking (Kim and Moon, 1998). In other studies on website design, Cyr et al. (2010) found that web-site color appeal is a significant determinant of website trust, and Lindgaard et al. (2011) also found strong correlations between visual appeal and trust in websites. A related variable to trust– *reputation* of an academic department – was correlated with aesthetics in a study of websites (Hartmann et al., 2007).

The effects of visual appeal on perceived *usability* was examined in several studies. Lee and Koubek (2010) found high positive correlations between usability and aesthetics before and after use. Like Tractinsky et al. (2000), they found that the effect of perceived aesthetics on perceived usability was stronger than the effect of objective performance on usability. Similar findings were obtained by Sauer and Sonderegger (2009). In another study, Sonderegger and Sauer (2010) found that participants using cell phones with high visual appeal rated them as more usable than participants using the unappealing devices. In a study of mobile phones, Quinn and Tran (2010) found that attractiveness accounted for as much variance in the SUS scores as effectiveness and efficiency. On the other hand, various studies found weaker or no such associations between visual aesthetics and usability (e.g., Lindgaard and Dudek, 2003, study 2; Hassenzahl, 2004a, Hassenzahl 2010; Thuring and Mahlke, 2007). The mixed findings suggest that the presumed association between perceptions of aesthetics and usability may not be universal. We elaborate on this point when we discuss the next category.

Visual aesthetics are considered a major force influencing perceptions of product *character* (Hassenzahl, 2003; Krippendorff, 2005) or *brand personality* (Park et al., 2005). It doesn't come as a surprise, then, that in a study on product choice, most participants mentioned aesthetic and symbolic roles most often as affecting product choice (Creusen and Schoormans, 2005). Still in online environments, Mandel and Johnson (2002) found hat color and image-based priming influenced online consumers' product choice. And Schmitt and Liu (2009) found that users are willing to sacrifice loading speed for a more aesthetically appealing webpage.

Following Norman's (Norman 2004) claim that "attractive things work better," perhaps the most intriguing question regarding the outcomes of visual aesthetic is whether it influences not only users' perceptions and evaluations of the system, but also their *Performance*. Recent studies have started looking for empirical evidence regarding this question (e.g., Moshagen et al., 2009). In a study of 11 data visualization techniques, Cawthon and Moere (2007) found positive relation between aesthetic data visualizations and performance of data retrieval tasks. Sonderegger and Sauer (2010) and Quinn and Tran (2010) similarly found more effective task performance when using attractive versus unattractive mobile phones. Van Schaik and Ling (2009), however, did not find relation between perceptions of classical and expressive aesthetics and performance measures.

Finally, it is important to note that visual aesthetics is considered a prominent antecedent of the concept of "User Experience" (Hassenzahl and Tractinsky, 2006; Sutcliffe, 2009; Law and Schaik, 2010). In a recent survey of the user experience (UX) literature, Bargas-Avila and Hornbaek (2011) found that emotions, enjoyment and aesthetics are the most frequently assessed dimensions of UX. Considering that aesthetics is also an antecedent of the other two dimensions, it appears that its role in influencing the UX is large indeed. To a large extent, it is related to almost all the ideas expressed in this chapter.

### 19.2.4 Moderating Variables

Linkages between perceived beauty and various outcomes or between design attributes and aesthetic perceptions, even if backed up by solid research evidence, common sense, or philosophical arguments, should not be considered universal or deterministic (Sutcliffe, 2010). First, as mentioned in the previous subsection, against studies that empirically found associations between aesthetic evaluations and evaluations of other perceived system attributes, such as usability, there are studies that found weaker or no such associations, indicating that at least under certain circumstances they do not hold. Second, in social settings, where research on the "beautiful is good" phenomenon accumulated evidence earlier and for much longer than in our field, findings suggest that the associations between attractiveness and perceptions of other human attributes are not unqualified (Eagly et al., 1991). Third, for all we know about socio-technical phenomena, it makes little sense that such deterministic relationships exist in a complex reality that involves individual, social and technological forces. Thus, adopting a contingent approach to the study of visual aesthetics would probably be more productive in describing if and how aesthetic evaluations mediate between various antecedents and consequences.

The challenge is then, to identify and examine how various factors serve to alter or moderate the aesthetic process. In Tractinsky (2006) I have provided a partial list of such potential moderators. The list included the type of system used (a typology that can span multiple dimensions such as consumer product vs. a computer application; small vs. large display; personal vs. public; hedonic vs. utilitarian, etc.), the use context (e.g., work vs. entertainment), cultural differences (national, sub-cultural, idiological), and so on. Individual differences constitute an interesting group of potential moderators, because people vary greatly in their sensitivity to aesthetic stimuli and in their aesthetic preferences (e.g., Bloch et al., 2003; Hoyer and Stokburger-Sauer, 2011). Jacobsen (2004) study found consistent intra-individual aesthetic judgments but strong inter-individual differences in beauty judgments. In addition, the group model of aesthetic judgment misrepresented about half of the study's participants. Pandir and Knight (2006) also found disagreement on aesthetic preferences in a study of different websites.

Contextual factors, such as domain and type of task are mentioned by Norman (2004) as important considerations for the type of aesthetic design required for users' performance and satisfaction. He argues that in certain domains (e.g., control rooms) attractive design may not necessarily be desired. Ben-Bassat et al (2006) found that people weighed more usability over aesthetic factors when faced with a performance-oriented task, and Van Schaik and Ling (2008) demonstrated that attractiveness ratings were affected by providing context for the evaluation task. In online shopping environments Cai and Xu (2011) found that the effect of expressive aesthetics on shopping enjoyment was stronger when shopping for hedonic products compared to utilitarian products.

Individual factors may also affect how anteceding variables (e.g., objective design attributes) are perceived differently by people with different aesthetic tastes (Hoyer and Stokburger-Sauer, 2011). In the domain of web-site design, Park et al., (2004) found that variability in user tastes is associated with aesthetic fidelity (i.e., the degree to which users felt the target impressions intended by designers). Individual differences were also found to affect the relative importance of aesthetics in people's preference of web-sites (Hartmann et al., 2008).

Attributes of the choice process were found to moderate the relation between aesthetic evaluation and product choice, especially when users are required to trade-off aesthetic for other system qualities. For example, Ben-Bassat et al (2006) found that system preference or choice were affected by aesthetics under ordinary conditions (e.g., questionnaires) but not when the participants had to bid for a system with which they will perform competitive tasks. Diefenbach and Hassenzahl (2007) showed that under a beauty-usability trade-off, although people may prefer more beautiful products to more usable ones, they choose the more usable product if they cannot justify choosing the more beautiful one.

Cross cultural studies have shown that national and professional cultures affect various relationships between aesthetic evaluations, their antecedents and their consequences. Several studies have demonstrated this moderating effect in the context of websites. For example, Cyr (2008) found effects of visual design on trust in China but not in Canada or Germany, and Cyr et al (2010) found different reactions to web-site color appeal in Canada, Germany and Japan. Hartmann et al (2007) found that the aesthetic evaluations and the importance of aesthetics are contingent on users' background (design vs. technical; Western vs. Asian).

The contingent nature of the aesthetic process is exemplified by Moshagen et al's finding that high visual aesthetics improved performance under poor usability but had no effect under high usability. Consequently, they quoted Liu's (Liu 2003) principle that ". . . ergo-aesthetic design does not imply that workplace or product designers should only use designs that are pleasing or attractive. On the contrary, ergo-aesthetic design advocates the careful and proper selection of aesthetic levels of design to fit the needs and characteristics of the intended use" (p. 1298).

## 19.3 FUTURE DIRECTIONS

The review above presented the results of empirical studies that examined antecedents and effects of visual aesthetics in HCI and various contingencies that moderate the relationship along this aesthetic process. In this section I would like to discuss several methodological issues and suggestions for future research in this area.

### 19.3.1 Methodological Issues

The review of research in the field uncovered several methodological issues that may also be involved in masking effects along the aesthetic process. One such aspect concerns evaluations that are more nuanced than overall aesthetic evaluations, which are quite common in the studies surveyed in this chapter. Studies that look for evaluations of aesthetic sub dimensions (e.g., Kim et al., 2003; Lavie and Tractinsky, 2004; Moshagen and Thielsch, 2010) can potentially yield richer accounts of the influence of design on aesthetic processes and on subsequent evaluations of the interactive system, attitudes towards it and interactions with it.

A related issue deals with the measurement of visual aesthetics evaluations or judgment. In the field of HCI, aesthetic evaluations were measured by a single item and by multiple-item scales. For example, Kurosu and Kashimura (1995) Tractinsky (1997), Schenkman and Jonsson (2000), Hassenzahl and Monk (2010), and Sonderegger and Sauer (2010) have used a single item asking about the beauty of the various applications and interactive products. Others, such as Schenkman and Jonsson (2000), Van der Heijden (2003) ,Moshagen et al., (2009) employed multiple-item scales to measure attractiveness. While multiple item scales are generally regarded as more reliable measures, single item scales have some practical advantages. In general, the main advantage if that single items make questionnaires shorter, reducing participants' fatigue and tendency to skip some of the items. In particular, the use of single items in the study of aesthetics allows quicker responses to stimuli in studies that focus on swift aesthetic responses (Lindgaard et al., 2006; Lindgaard et al., 2011; Tractinsky et al., 2006). The tension between scientific directives and practical constraints may not be as severe as it first appears. Studies suggest that when dealing with a concrete object (e.g., the application or product to be evaluated) and a concrete attribute of the object then single item measures are as valid as multiple-item scales (Gardner et al., 1998; Bergkvist and Rossiter, 2007). While the scientific community may have a hard time defining what is meant by the concepts of "aesthetics" or "beauty" – perhaps due to the multiple disciplines that deal with these concepts and which attach different meanings to them, my experience is that ordinary people's intuitive interpretation of the terms correspond closely to the dictionary definition provided above, which guides research on visual beauty in HCI. This point may be worth further research for corroboration, but if correct, future scientific and practical studies would be able to safely use single item measures of visual aesthetics.

Research on visual aesthetics in HCI has employed a mix of experimental and correlational designs. Because some of the most interesting aspects of visual aesthetics research involve questions of cause and effect, experimental studies would appear to provide the most conclusive evidence. It is straightforward to study basic and relatively simple design effects (e.g., symmetry using basic patterns) on aesthetic perceptions using experimental designs (Bauerly and Liu, 2006; Winkielman et al., 2006). However, it becomes increasingly more difficult if we want to study the effects of aesthetic design using more complex and ecologically valid stimuli, like those used in correlational studies (e.g., Lindgaard et al., 2006; Hassenzahl and Monk, 2010). Thus, employing experimental designs using elaborated and realistic stimuli is a major challenge. Ideally, to test causal effects studies would manipulate design attributes independently of each other to separate aesthetic perceptions from perceptions of other system attributes. In practice, however, this is very difficult to accomplish due to the a priori association of these attributes (Moshagen et al., 2009). One frequent consequence of attempting to achieve this independent aesthetic manipulation is that it creates a relatively small variance in the manipulated stimuli (otherwise, strong aesthetic manipulations might also cause differences in other experimental factors). The danger is that small variance and the lack of strong aesthetic condition would in turn weaken the effects of visual aesthetics. Another challenge in manipulating or selecting aesthetic stimuli in experimental designs relate to whether the degree of aesthetic stimuli is defined “on average” (e.g., by a pilot study or manipulation check) or is defined separately for each individually (e.g., in a procedure described by Tractinsky et al., 2000). The advantage of the latter approach is improved probability that individuals who are assigned to various aesthetic groups in the experiment indeed perceived the stimulus in a way that corresponds to their group (as opposed to a stimulus that may belong to that group on average, but which doesn’t match the participant’s aesthetic taste). This would increase the effect size of the aesthetic manipulation. On the other hand, such a procedure usually requires pre-experimental exposure to a set of potential aesthetic stimuli. This process may later interact with the experiment (e.g., by creating expectations towards the experiment), and may create undesirable noise.

### 19.3.2 Future Research

This chapter has reaffirmed that visual aesthetics is associated with a range of HCI-related variables. However, it is also apparent that our understanding of the contingent nature of the processes that surround visual aesthetics is still limited. Thus, further exploring these contingencies (i.e., the conditions under which perceptions of visual aesthetics or its effects change due to contextual factors) appears to be more beneficial to the advancement of knowledge in the field than attempting to confirm direct relationships along the visual aesthetics process chain.

Most studies to date have concentrated on people’s first reactions to visual aesthetics or to short term impact of aesthetic design. Studies are also characterized by providing participants a limited set of aesthetic stimuli to choose from. The problem with such sets is that they do not necessarily include designs that are viewed by the participants as beautiful. In addition, such studies rarely represent reflective aesthetic value to individual participants. Such stimuli may be adequate for creating short term impressions, but they are hardly adequate for assessing contemplative evaluations and longer term evolvement of aesthetic processes. Thus, to expand the picture of visual aesthetics in HCI, future research should emphasize more reflective evaluation and contemplation of designed products and environments.

Another research topic that has yet to receive attention is the (dis)connect between designers and users. In other design disciplines, studies have found significant differences in aesthetic evaluation between laypeople and designers (e.g., Nasar, 1997 and Gifford et al, 2000, in the field of architecture). In HCI such differences were found by Korman-Golander (2011) between designers and software engineering students in assessments of web-site design trends. Similarly, Inbar et al. (2007) and Bateman et al. (2010) found that the minimalist design recommendations for charts made by Tufte's (1983) influential critique of "chartjunk" practices do not resonate with people's actual preference of chart types. To date, I am aware of only a few studies (e.g., Park et al. (2004) and Bateman et al, 2010) that have tried to tease out the sources of those differences, and to offer methods that would help bridge the gap between designers and other members of the development team and between the development team and intended users.

In his seminal work on the extended self, Belk (1988) listed various product categories in which there is significant image congruity between a brand or a product category and self images of owners. The list does not include IT products, but there are good reasons to expect that such congruity holds, for example, in the choice of personal computing, smart phones, media players, software, etc. We may then explore what role visual aesthetics plays in motivating people to choose those interactive media.

In discussing our early work on the relationships between aesthetic and perceived attributes of the system we called on researchers "to shed more light on the cognitive and/or affective processes that lead users to associate interface aesthetics with other system attributes" (Tractinsky et al., 2000, p. 140). Several studies have recently taken on this challenge. For example, Hassenzahl and Monk (2010) suggest that perceived aesthetics affects users' evaluation of the system's goodness, which in turn influences evaluations of the system's usability. Similarly, Lindgaard et al. (2011) suggest that the initial attraction generated by a system's aesthetics forms "a general attitude" of aesthetic, which is later refined through further use of the system and reflection based on high level emotional and cognitive processing. However, it seems that there is ample room for continuous research on the mechanisms that underlie these relationships. In particular, studies about the interplay of emotional and cognitive factors (Sun and Zhang, 2006; Thuring and Mahlke, 2007) at the three levels of processing (Norman, 2004) are sorely needed.

Studies of visual aesthetics in HCI have for the most part concentrated on the relatively stable properties of the user interface design. Thus, studies have used website screenshots, interactive products' hardware design, or general aesthetic features of systems. Little attention was paid to dynamic aspects of visual aesthetics. With the increased embedding of dynamic visualizations , video clips and various animations in interactive systems we need to have a better grasp of their aesthetic qualities (Chen, 2005). Some initial steps in this direction can be found in a study of perceived aesthetic dimensions of animations, done in the context of in-car presentation of eco-driving information (Tractinsky et al., 2011b)

Finally, much of the variability in people's assessment of visual design and the effects of visual design can be attributed to individual and cultural factors. These factors may include differences in sensitivity to visual aesthetics, different weighing of visual aesthetics when appraising systems and products, and different notions of what is considered beautiful. Such studies could explore why and how people personalize interactive systems and products (e.g., Tractinsky and Lavie, 2002, Tractinsky and Zmiri, 2006); why some people prefer ornamented charts or web pages while others prefer minimalist styles (e.g., Inbar et al, 2007; Bateman et al., 2010); why reactions to website color treatments differ among different cultures (Cyr et al., 2010); and whether people belonging to different trend and fashion adoption groups prefer different website designs (Korman-Golander 2011).

## 19.4 CONCLUSION

Interest in visual aesthetics in HCI has grown considerably over the last 15 years. From a short conference paper that reported correlations between perceived aesthetics and apparent usability (Korosu and Kashimura, 1995) to a rich field of inquiry. It is possible that the interest in the field was motivated by provocative titles such as "What is beautiful is usable" (Tractinsky et al., 2000) and "Attractive things work better" (Norman, 2004). It is more likely, however, that it corresponded to technological and societal changes that have swept our lives over that time and reshaped the field of human-computer interaction.

## 19.5 COMMENTARY BY JEFFREY BARDZELL

How to cite this commentary in your report

### Jeffrey Bardzell



Jeffrey Bardzell is an Associate Professor of HCI/Design and new media at the School of Informatics in Indiana University - Bloomington. With a Ph.D. in Comparative Literature and Minor in Philosophy, Bardzell brings a humanist perspective to HCI and is known for developing a theory of interaction criticism. His other HCI specialties include aesthetic interaction, user experience design,...

Jeffrey Bardzell
Jeffrey Bardzell is a member of The Interaction Design Foundation

> “The Mathemagician: “You couldn’t have tea for two without the two, could you?”
>
> King Azaz: “You couldn’t have tea for two without the tea, could you?””
>
> -- *The Phantom Tollbooth*

The thesis of my commentary is that Noam Tractinsky’s chapter on Visual Aesthetics reflects and champions only a portion of work on aesthetics that has influenced HCI, that he has (hopefully unintentionally) marginalized alternative approaches, and that a more balanced picture needs to be offered. To defend this thesis, my commentary will do the following:

- Demonstrate through close reading that Tractinsky’s chapter offers a philosophical theory of aesthetics and also denies that it does so as part of a rhetorical strategy to avoid engaging with aesthetic theory
- Propose three arguments motivating visual aesthetics in interaction design

- Critically evaluate Tractinsky's (what I will call) "aesthetic processing theory for HCI" in light of these three arguments
- Explore rival aesthetic theories from the humanities
- Introduce HCI research and design that leverages these rival theories and evaluate them in light of these three arguments
- Argue that we need a more balanced and comprehensive view than the one Tractinsky offers if the interaction design community is to address all three arguments motivating visual aesthetics and interaction design

The most substantial criticism of Tractinsky's article that I will make is that his account of aesthetics in general and visual aesthetics in HCI in particular is extremely limited, rather than comprehensive as he promises, and it excludes both major aesthetic ideas and also major aesthetic contributions to interaction design. Such a marginalization therefore offers, in my view, a distorted account of his putative topic—visual aesthetics in HCI—and might encourage readers to miss opportunities to work towards a goal that *all* of us have in common: a desire to make interaction more aesthetic.

While my commentary takes a critical position with regard to Tractinsky's essay, I want to stress up front both that Tractinsky's research, as well the work of those within his tradition (including nearly all my fellow commenters on this chapter), has had enormous positive influence over the years in HCI, that I myself both teach and use such work, and that I broadly agree with Tractinsky's prescriptions for the future of such research. My purpose is not to attack what clearly is a rigorous, useful, and influential research approach; rather, it is to critique its positioning and its limits and to explore alternative formulations that complement and strengthen, but do not replace, its place in HCI.

At the end of the day, we are successful if we have helped designers make interactions that are more aesthetic, not if we win academic turf wars.

### 19.5.1 An anti-theory theory and its consequences

Tractinsky opens his chapter by defining and scoping his operational understanding of "visual aesthetics" by separately defining "visual" and "aesthetics":

> "I use the term "aesthetics" in its fairly ordinary and common sense as reflected in dictionary definitions such as "an artistically beautiful or pleasing appearance" (The American Heritage Dictionary of the English Language), or as "a pleasing appearance or effect: Beauty" (Merriam-Webster's Collegiate Dictionary). The term "visual" indicates concentration on the visual sense, which is the central human sense, occupying "almost half the brain" (Ware, 2008, ix). Thus, this chapter is not about various other phenomena studied under the "aesthetics" heading, such as literary aesthetics, abstract forms of aesthetic experiences or criteria.... or reactions to object qualities that do not immediately and primarily stem from its visual attributes"
>
> *-- Section 19.1*

By using two standard dictionary definitions and by bracketing aside the "various other phenomena studied under the 'aesthetics' heading," Tractinsky categorically disengages with millennia of aesthetic thinking in philosophy, art history, literature, architecture, and film. Yet by referencing brain sciences in his definition of the visual, he indicates his willingness to engage with scientific scholarship. Thus, Tractinsky has signaled his intention—one that he will carry out throughout the article—to be scholarly about the empirical science of aesthetics and strategically unscholarly about philosophy of aesthetics. The latter is not an accusation from me but rather how he describes himself:

> “ While the scientific community may have a hard time defining what is meant by the concepts of “aesthetics” or “beauty”—perhaps due to the multiple disciplines that deal with these concepts and which attach different meanings to them, my experience is that ordinary people’s intuitive interpretation of the terms correspond closely to the dictionary definition [sic] provided above, which guides research on visual beauty in HCI. ”

-- *Section 19.3.1*

Here we have the strong claim that interdisciplinary attempts at defining aesthetics confuse the scientific community, and that the way forward is deceptively simple: to base aesthetic research in HCI on what he calls “ordinary people’s intuitive interpretation of the terms,” which are reflected in the dictionary definitions and to disregard all that multidisciplinary handwringing. With that established, research is “primarily empirical and ... characteristically descriptive (i.e., ‘what *is* considered beautiful’) rather than normative (i.e., what *should* be considered ‘beautiful’)” Section 19.1. Here, Tractinsky is again positioning himself and other researchers *outside* of aesthetic debates: his job is simply to *discover* what is already out there in the world and not to take positions on (which would place him inside) aesthetic debates.

Thus, Tractinsky’s success with his “ordinary language” definition of aesthetics hinges whether the view of aesthetics that he presents reasonably reflects people’s ordinary views of aesthetics. If it does, then his categorical rejection of millennia of humanist scholarship on aesthetics seems reasonable, since it does not appear to provide anything useful for his research project and may even harm it through multiple and confusing technical definitions. Moreover, if he does hold to an ordinary definition of aesthetics, then he can justifiably argue that he is outside of aesthetic debates and simply discovering what's already there (which is what he means when he talks about being "descriptive" and not "normative").

But if Tractinsky's work turns out *not* to be based on ordinary views of aesthetics, then he has set up two problems for himself: first, he is vulnerable to the criticism that his approach lacks scholarly rigor because he seems simply to have chickened out of engaging with the conceptual difficulty that most of the rest of us find intrinsic to aesthetic reasoning; and second, he loses this right to claim that he is merely descriptive and not normative, and therefore that his proposed science of aesthetics effectively becomes another candidate philosophical view of aesthetics, and hence subject to the very sort of philosophical critique that he seeks to circumvent by avoiding theory in the first place.

I argue that it is easy to see that the notions of aesthetics that Tractinsky cites with approval over the course of the article neither match the brief and nearly vacuous dictionary definitions he quotes, nor does the view of aesthetics that he promotes reflect the common, non-scholarly, intuitive views of aesthetics held by ordinary people. Therefore, I argue that what "guides research on visual beauty in HCI" is not a simple idea intuitively shared by most people, but rather is a sophisticated, technical, and robustly academic theory, that this theory inevitably has normative dimensions to it (and, incidentally, it's hard for me to understand how anything that has "implications for design" is not intrinsically normative), and therefore deserves critical scrutiny in order to be used rationally. In short, I argue that Tractinsky introduces an aesthetic dogma that is cloaked in supposedly descriptive empirical science; as a field, we need to disentangle the two so that we don't throw out the baby with the bathwater.

### 19.5.2 Tractinsky's extra-ordinary language definition of aesthetics

In the course of the article, Tractinsky cites with approval a diverse (and unabashedly scholarly) array of aesthetics concepts that anyone would be hard-pressed to claim are part of what he calls "ordinary people's intuitive interpretation of the terms" and certainly won't be found in collegiate dictionaries under "aesthetics." These include the following:

- The three Vitruvian principles of architecture (*firmitas*, *utilitas*, *venustas*) are offered as an ancient means of articulating different dimensions of design value, which establishes an analytic understanding of design value comprising a structural relationship among strength, durability, and structure; usefulness and suitability; and beauty (Section 19.1.2.1).
- Gestalt psychology of perception is used to explain why usability and beauty are harmonious, rather than conflicting values (Section 19.1.2.2).
- Tractinsky's own influential use of "classical" versus "expressive" aesthetic dimensions is offered as a further means of exploring relationships among usability and beauty (Section 19.1.2.2).
- Psychological understandings of aesthetics, too numerous to mention here in full, are offered, including the idea that "the value of visual aesthetics" comes from "pleasure and wellbeing," "basic human need," and "perhaps ... evolutionary processes" (Section 19.1.3.1), that "aesthetic experience" is a combination of "affective responses and reflective thought" (Section 19.1.3.1), that aesthetics has something to do with the psychology of the self (Section 19.1.3.2), and that "aesthetic stimuli" cause "very fast"

"aesthetic impressions" (Section 19.1.3.3), and above all that aesthetics can be modeled as using an information processing metaphor (Section 19.2).

- Design understandings of aesthetics are summarized as respecting audience, being sensitive to needs and desires, and designing with effort and care (Section 19.1.4.3).
- Norman's division of "aesthetic perceptions and evaluations" across "visceral, behavioral, reflective" ... "levels" ... "of processing" is summarized and championed several times (e.g., Section 19.2.2).

There is no space here to evaluate these different concepts (each of which brings with it insights and difficulties), and Tractinsky is certainly justified in outlining them as influential and important aesthetic ideas that have been explored in psychology, design, and more recently HCI. But in embracing all of these ideas, Tractinsky has outlined a philosophical infrastructure for an academic theory of aesthetics and departed from a commonsense or dictionary notion of the aesthetic. I believe I can reasonably assert that none of the ideas in the list above are part of "ordinary people's intuitive interpretation of the terms" *and* I can also reasonably assert that all of these ideas are foundational to Tractinsky's research project. Therefore, his anti-theory stance is invalid: he has constructed for himself a theoretical apparatus constituted by a set of technical and interlocking ideas, and he is not relying on a simple dictionary definition as he claims.

And it is building from this apparatus that Tractinsky offers a particular academic theory of aesthetics in Section 19.2, where he outlines a flow model for what he calls "the aesthetic process." According to this model, *design variables* (including "low-level" attributes such as use of color and symmetry and "high level" attributes such as novelty, typicality, and fluency) lead to (cause?) *aesthetic valuations* (such as visceral, behavioral, and reflective ones, following Norman) and in turn aesthetic valuations lead to (cause?) *outcomes* (such as affects and emotions, brand trust, perceived usability, perceived product character). Each of these relationships is *modified* by the type of system used, cultural inputs, domain, type of task, and aesthetic tastes. In this model, Tractinsky intermingles **traditional aesthetic categories** (e.g., symmetry, form, composition, balance of typicality and novelty, diversity, craftsmanship, expressivity) and **the language of experimental psychology** (e.g., the information processing metaphor, variables, moderators, input/output, performance, motivation).

This is a very sophisticated theory, and at this point I would like revisit Tractinsky's foundational claims:

1. The two dictionary definitions (which both assert that beauty is that which is pleasing) *accurately reflect* ordinary people's intuitive understandings of aesthetics *and these are* that which "guides research on visual beauty in HCI"
2. And therefore:
   A. No academic theory of beauty beyond a dictionary definition is needed to pursue this research *and*
   B. Empirical science on visual aesthetics for HCI is descriptive and not normative because it stays out of multidisciplinary aesthetic debates

But merely summarizing Tractinsky shows that what he has offered here is far more than the dictionary definition and adds up to *an information processing theory of aesthetics* in which design inputs yield evaluation outputs, and evaluation outputs become inputs for outcomes outputs. According to this theory, the process itself has "moderators" including use context, system attributes, culture, and individual differences as inputs. This is substantially more specific and more guiding than his dictionary definition ("a pleasing appearance or effect: Beauty"), a phrase that is vague to the point of meaninglessness and is obviously insufficient to guide the empirical research of aesthetics in HCI!

I have, I believe, established that in spite of what he claims, Tractinsky operates with a much more sophisticated aesthetic philosophy than a dictionary definition, but I have not commented on how his theory relates to ordinary people's intuitive understandings. But I think some very simple reflection can disabuse us of that pretention as well. For example, I teach a course called *Interaction Culture* to HCI students within a School of Informatics and Computing. In the opening minutes of the first day of that class, before I even introduce myself or give students the syllabus, I show the first few minutes of an art film that has achieved some popular success. This year, I showed the first 5-6 minutes of *Run Lola Run*, a 1998 German action film with a philosophical subtext. The opening of the film includes 3D computer graphics, 2D cartoon animation, 2.5 post-production compositing (e.g., titles), live motion acting, heavy image manipulation, and an intense techno beat; in addition to the bewildering visual assault there is also a short but mysteriously philosophical verbal script whose relationship to the images is not obvious. After showing this introduction once, I ask students to simply talk out loud about their reactions to it. Some describe *how it made them feel*—excited, anxious, curious. Others talk about *symbols that constitute the work as an artifact*—how the heavy use of clock imagery and the metronomic beat of the techno soundtrack reinforced each other. Others offer suggestions about *what the director may have intended* or was trying to say or do, how this fits in the German cinematic tradition, etc. Still others talk about *what was happening when the film was made* (the 1990s in Germany or popular culture in general). Are these not ordinary, common, and intuitive aesthetic reactions? Such interpretative strategies—and not aesthetic processing theory—are taught to us as children in schools and at home and come almost naturally to us as adults. My sense is that if one really wants to understand what ordinary people intuitively do, all one needs to do is watch ordinary people intuitively encounter beautiful things. Tractinsky goes far beyond that in his research, and rightly so, but it is disingenuous to claim that he neither needs nor uses any disciplined academic theory of aesthetics.


Everything else is pure theory.






LOLA

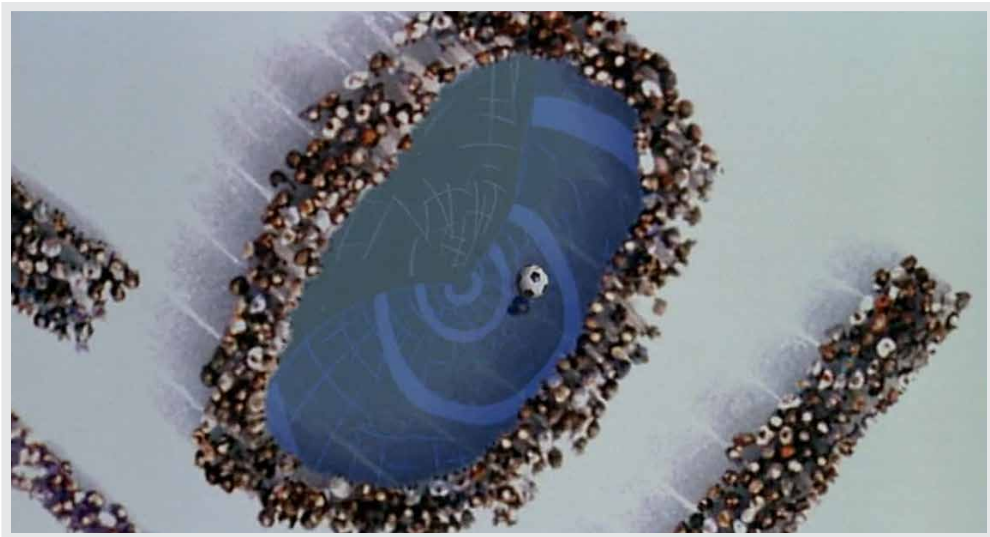






Figure 19.1: A sequence from the credits of Run Lola Run.

So Tractinsky's aesthetic processing theory is *not* intuitive or common. Nor is it an empirically discovered fact in the world. That theory is thus a *philosophical theory of aesthetics*, or, in the language of logical positivism (which I personally reject but am not certain that Tractinsky does), a *dogma*. Aesthetic processing is a theory of aesthetic response constructed out of the methodological and conceptual apparatus of information processing psychology and adapted using aesthetic vocabulary from both the sciences and the arts. For example, Tractinsky's own seminal distinction between "classical" and "expressive" aesthetics are both derived from the history and philosophy of art, the former attributable to philosophers such as Hutcheson, Bell, and Beardsley and the latter attributable to philosophers such as Langer and Collingwood. (Both concepts have also been *developed* and *critiqued* for over a century, and the conceptual difficulties of each are well known among analytic philosophers of art, if not the HCI community.)

### 19.5.3 The aesthetic processing theory and its discontents

My argument thus far has been largely philosophical, seeking to show that the conceptual edifice on which Tractinsky builds his aesthetic processing theory is flawed inasmuch as it claims to be a-theoretical when it clearly is not.

However, the real point of all this is much more practical: I want to show that the theoretical blindness built into Tractinsky's philosophy of visual aesthetics has important consequences for HCI that need to be dealt with. Two of them are as follows:

- The a-theoretic position exempts itself from critical scrutiny, since it denies the existence of its own theoretical constructedness and normative commitments, presenting itself innocently as mere empirical data; such data, this position implies, can be scientifically but not philosophically interrogated.
- The a-theoretic position is used to marginalize Tractinsky's rivals—basically, anyone with a humanist or openly theoretical orientation

to aesthetics who seeks to contribute to HCI; significantly, all such research, and in spite of the fact that some of it has been extremely influential in the field, has been all but completely ignored in Tractinsky's essay (and those of the other commentators).

Of the two, the second is a more serious shortcoming, especially given that Tractinsky claims that "The objective of this paper is to survey the field of visual aesthetics in HCI," when in fact it surveys only a favored subset of that field. I will address both of these practical consequences in what follows.

### 19.5.4 What do we want from visual aesthetics in HCI?

Scientific research is expensive, and one way or another the public pays for it, and so any scientific agenda should deliver some sort of public good. What is the social value of aesthetic research in HCI? Let us follow Tractinsky's example by beginning with the dictionaries:

> " an artistically beautiful or pleasing appearance" (The American Heritage Dictionary of the English Language), or as "a pleasing appearance or effect: Beauty "

These definitions are not terribly helpful. The only public good I can imagine deriving from them is that this research will make interacting with digital systems more "pleasing." But surely this is a weak argument: just as my insurance company won't pay for voluntary cosmetic surgery to make me look more pleasing, I can't imagine policymakers in this era of austerity investing in scientific research to make user interfaces more "pleasing."

Tractinsky himself offers numerous and much better arguments in the course of his essay. He notes that aesthetics has long been integrated within design disciplines whose professional and socioeconomic success is beyond dispute and whose theories and methods can be leveraged in HCI and interaction design. He notes that Gestalt psychology has shown that aesthetic criteria are linked with other design values, including usefulness and suitability, his most powerful argument to the HCI community, which historically has had a orientation towards the useful. He adds that aesthetics satisfies human needs (not merely superficial desires), contributes to wellness, and seems to be linked to the formation and experience of the self, making the argument that aesthetics is good for people's lives. He also notes that aesthetics helps otherwise similar products differentiate themselves, thus contributing economic value (which, in the case of Apple, has been substantial). I accept each of these (normative) arguments as stated and stress that he has here offered a number of social benefits that can emerge from this research, most of which are functional in nature: aesthetics supports usability, aesthetics satisfies needs, aesthetics contributes to the self, aesthetics contributes to economic prosperity.

If we turn to other philosophers of aesthetics besides Tractinsky (and I count him as one, whether or not he does), we can see many other arguments commonly made that would support the idea that this research contributes to the public good. Common claims in the aesthetic literature include the following statements about aesthetic response and/or aesthetic experience (synthesized from Bardzell, 2011):

- Aesthetic experience is intellectually and emotionally rich and fulfilling, thus improving quality of everyday life. In HCI, McCarthy & Wright (2004) build on the aesthetics of philosopher John Dewey to propose a holistic view of good experience, so that experience designers have something to orient their work toward.
- It can educate our perception and challenge and develop our cognitive abilities (e.g., reasoning, sense-making, learning) in worth-

while ways. An emphasis on active, rather than passive computer use has long been advocated in the work of HCI researcher Yvonne Rogers (2006), and while she doesn't invoke the language of aesthetics, she clearly is thinking along comparable lines.

- It contributes directly to human knowledge and understanding of the world. Researchers in critical design (Dunne & Raby, 2001) have used aesthetic designs to generate knowledge for and about interaction design.
- It can be individually enlightening and ethically uplifting, e.g., by heightening one's capacity for empathy. Critical design researchers have also argued that their methodologies contribute to these outcomes.

Going back to Plato, aesthetics has been implicated not only in pleasures but also its role in contributing to (or detracting from) an educated and responsible public, and this predisposition is amply reflected in the list above. As interactive technologies continue to replace older media forms in mediating how people interact with themselves, each other, and the world, making interaction aesthetic in these senses seems to be imperative, rather than optional. The cultivation through aesthetic engagement of ourselves as perceptive, imaginative, and insightful citizens (an epistemological position) would seem to depend increasingly on human-computer interaction.

I have briefly sketched 3 simple arguments justifying aesthetic interaction: a hedonic argument, a functionalist argument, and an epistemological argument. While I personally support all three, it seems the second (Tractinsky's functionalist argument) and the third (aesthetic philosophers' epistemological argument) could be the most compelling for policymakers as well as researchers and practitioners within the field of HCI and interaction design.

### 19.5.5 Critically assessing the aesthetic processing theory in HCI

Now that I have sketched out three primary arguments in favor of pursuing visual aesthetics in HCI, and I have earlier established that Tractinsky offers a theory of aesthetic processing as a means to do so, so we are finally in a position to evaluate the strengths and weaknesses of this aesthetic processing theory for HCI.

#### 19.5.5.1 Strengths of the aesthetic processing theory for HCI

Tractinsky does much of my work for me here, since his article systematically summarizes the achievements of this tradition, and by and large I accept his account of that at face value; as I said in my introduction, I use aesthetic processing research in both my own research and teaching. I'll add a few points here for emphasis.

I begin with the point that although aesthetic philosophy, literary theory, and art history (etc.) have anticipated and expounded on many of the concepts used in aesthetic processing theory, nothing that the humanities tradition has done anticipated the exciting and specific findings about how quickly judgments and evaluations are made, how such judgments are causally linked to closely related phenomena (e.g., affect, human needs), and how all of the above influence behavior. If one accepts Norman's three-part distinction between visceral, behavioral, and reflective aesthetic perception and evaluation (and I do not, but that's beside the point here), it is clear that the aesthetic processing approach has offered unparalleled insight into the visceral processing—a fact that is not lost on marketers as well as designers. Some of the questions that this research is well positioned to answer include user perceptions of a system's usability and beauty, behavioral consequences of aesthetic valuations, attitudes (such as trust) towards organizations represented in systems, brand personality, and perceived system performance.

Aesthetic processing research provides new discoveries about the very mechanisms of aesthetic perception and experience, and their implications go beyond

HCI and should influence anyone interested in visual aesthetics in the humanities and sciences. In their critique of McCarthy & Wright for failing to clarify much specifically about the *content* of experience, Hassenzahl, Diefenbach, & Göritz (who in my view operate in a similar paradigm with Tractinsky) argue that this sort of approach can reveal much about how experience is constructed from perceptions, motivations, affect, and judgments and thereby offer useful implications for designers (Hassenzahl, Diefenbach, & Göritz, 2010); they are right about the strength of their approach compared to McCarthy & Wright's, though I believe McCarthy & Wright also offer complementary strengths that Hassenzahl, Diefenbach, & Göritz do not acknowledge. Anyway, aesthetic processing approaches to aesthetics have revealed much about the nature of fast aesthetic judgments of interactive systems, have done so with useful implications for design, and have offered compelling evidence to support their findings.

Another benefit of this approach is that Tractinsky's work in the 1990s decisively undermined prevailing attitudes (especially the high-profile urgings of Norman and Nielsen at the time) that viewed the aesthetic as inherently in conflict with the usable. Tractinsky helped change the field by offering evidence that usability and aesthetics were not, in fact, in conflict. In so doing, he helped create space for others of us interested in aesthetic interaction, not by making a nice argument about aesthetics, e.g., using Dewey or Heidegger to argue for a more robust aesthetic sensibility, but by proving the prevailing wisdom wrong on its own terms: experimental science. I am hardly a fan of scientism for its own sake, but the ability of this paradigm of research to leverage science to contribute to the scholarship of aesthetics (and aesthetic philosophers and literary theorists *are* beginning to read cognitive science on aesthetics), and specifically to advocate successfully for more work on aesthetics in HCI, is an enduring achievement.

Of the three arguments I offered supporting research on visual aesthetics in HCI (hedonic, functionalist, and epistemological), a strength of this tradition is that it speaks to the first two in powerful ways.

#### 19.5.5.2 Weaknesses of the aesthetic processing theory for HCI

As with any synoptic theory of aesthetics, the aesthetic processing model as presented has a number of practical weaknesses.

The first weakness stems from the fact that the aesthetic processing model differs in important ways from our common experience of the aesthetic: as I have argued, it is too steeped in information processing theory to fit with an ordinary person's experience. One aspect of that is the reductive approach that aesthetic processing theory relies on. Its level of analysis is often faster than what humans are consciously aware of: for example, the finding that broad aesthetic judgments are made within *x* number of milliseconds may have application for professional designers, but it certainly sits outside of an ordinary person's experience. Likewise, the reduction of holistically experienced phenomena (e.g., emotion) into constituent, measurable parts is also alien to the common aesthetic experience. Finally, many of the example interfaces shown in these studies are frankly ugly (see Figure 19.3 in the main chapter, and Figures 19.1A-B in Lindgaard's comment for typical examples), and the reason for this, I believe, is that the researchers' intellectual goal tends to be something like "given that we have to design, e.g., a Windows email dialog box, what choices will make it be perceived as more rather than less beautiful?" rather than "how do we design a beautiful interaction?" It is a strength of aesthetic processing theory that it is able to answer the first question, but it is a weakness of the approach that it can't put Humpty Dumpty back together again, that is, that once complex wholes are reduced into parts, the process can't be reversed. Pragmatically, this research is not typically used to promote beautiful interactions, but rather *more* beautiful (than they otherwise would have been) interactions.

Second, the aesthetic processing model is also comparatively weak at ethical and socio-cultural considerations. I agree with the ancient Greeks in seeing ethics and aesthetics as so deeply intertwined as to be inseparable, but in doing so, one must move the aesthetic from the realm of the perceptual and into the hermeneutic. Tractinsky writes, "this chapter is *not* about ... our reactions to object qualities that do not immediately and primarily stem from its visual attributes" (Section 19.1), so he seems to be excluding attributes that influence how we perceive the visual in the first place. In the Strengths section, I mentioned some of the specific contributions of empirical approaches based on the aesthetic processing theory: user perceptions of a system's usability and beauty, behavioral consequences of aesthetic valuations, attitudes (such as trust) towards organizations represented in systems, brand personality, and perceived system performance. Not only do none of these have a strong ethical dimension to them, but worse, armed with the findings of this research, marketers and designers are in a better position to *manipulate* users, because so much of this research provides practical guidance on how viscerally to influence perceptions, behaviors, and affects through design choices.

Aesthetic philosophers have long taken to task theories that, like this, focus tightly on the perception of objective visual qualities precisely because of concerns about manipulation. For example, analytic philosopher Mary Devereaux (1998) investigates this issue in Riefenstahl's *Triumph of the Will*, a Nazi propaganda film that was made with such stunning cinematic vision and craftsmanship that it won awards across Europe in the late 1930s and continues to influence the language of cinema today. If we limit ourselves to "objects or qualities" that "immediately and primarily stem from its visual attributes" we have little mechanism for dealing with the seductive evil of Riefenstahl's film; on the aesthetic processing theory, such a film is beautiful. Now, *Triumph of the Will* is an extreme case, but it calls into question the hermeneutic effects of moral sensibility within of aesthetic sensemaking and reflection. Consider a 2007 video game produced and distributed by Hezbollah called *Special Force 2*, a first-person shooter in which the player's goal is to shoot and kill as many Israeli soldiers as possible: ethics and aesthetics will converge in human-computer interaction in diverse and socially important ways.

Courtesy of Bundesarchiv, Bild 102-04062A. Copyright: CC-Att-SA-3 (Creative Commons Attribution-ShareAlike 3.0).


*Courtesy of Bundesarchiv, Bild 183-K0326-0503-003.. Copyright: CC-Att-SA-3 (Creative Commons Attribution-ShareAlike 3.0).*


*Courtesy of Bundesarchiv, Bild 102-04051A. Copyright: CC-Att-SA-3 (Creative Commons Attribution-ShareAlike 3.0).*

Figure 19.2 A-B-C: The pageantry of the Nuremberg rally was partly constructed in order to be visually pleasing on film.

An example pertaining to the social context of aesthetic response is Duchamp's *Fountain*, which is an otherwise ordinary urinal that the artist signed (under a *nom de plume*) and put on display in a museum, making it "art." There is nothing visually interesting about this urinal: if it is art, it is only so on account of its having been placed in an art museum, on a pedestal as a work of art, by someone recognized by his community as an artist. All of the aesthetic processing involved in this case is at best loosely connected with visual stimuli in the sense that Tractinsky uses the term. In his Commentary to this chapter, Marc Hassenzahl touches on this issue, noting that "beauty" is socially constructed rather than individually processed, though I note that Hassenzahl's idea has already been developed and debated for decades in philosophical aesthetics, yielding Arthur Danto's influential notion of "artworld" and three versions of Georgie Dickie's "institutional theory of art." One risk of the a-theoretical position outlined by Tractinsky is that it runs the risk of reinventing wheels.


FIGURE 19.3: Duchamp's 'Fountain' deliberately challenged traditional aesthetic notions of beauty and artistic achievement.

A third weakness is the lack of medium specificity in Tractinsky's account. An important topic in many aesthetic fields is an effort to discover and/or articulate what a given medium is uniquely good at presenting for aesthetic attention: what can film do that painting cannot, or vice-versa? The aesthetic processing model applies equally to all visual forms, painting, (visual) HCI, film, sculpture, etc. But given that we are in HCI and interaction design, it would seem that *interactive*, rather than *visual*, aesthetics would be the target. This concern is raised by Jinwoo Kim is his Commentary to this chapter. This weakness can presumably be explained by the fact that the aesthetic processing model, by virtue of being a *processing* model, is fundamentally about the cognitive (i.e., processes internal to an individual body and its cognitive processing) and thus has less to say about the visual artifact itself, except inasmuch as it is perceived and rendered available to consciousness.

The final weakness I will mention is that the aesthetic processing theory struggles to deal with the notion of skilled or expert interpretation, e.g., the idea that a critic or a designer has a more robust or better understanding than a layperson—a claim that both critics and designers make, justifiably in my view, because such an understanding is the foundation of their professions and what they are trained for. Tractinsky himself acknowledges this problem in Section 19.3.2 when he calls for more work regarding "the disconnect between designers and users," but he does not acknowledge the design research (e.g., Schön, 1983; Buchanan, 1995; Cross, 2007) or HCI research (e.g., Lowgren & Stolterman, 2004; Greenberg & Buxton, 2008; Bardzell, 2009; Bardzell, 2011) that has already explored this issue. Lacking a solid account of how legitimate subjective expert judgments are formed, exacerbated by a rejection of the normative aesthetics that they imply, the aesthetic processing account leaves little epistemological space for designers of aesthetic interactions to be anything other than ordinary people armed with empirical data—a characterization that is far from the self-perceptions of most designers and is all but senseless when applied to artists.

In short, the weaknesses of the aesthetic processing account is that it sheds little insight on the third argument in favor of visual aesthetics for HCI I described above: the epistemological one, which focuses on how aesthetic encounters have the long term effect of cultivating our capacities for imaginative perception, insight, critical thought, and empathy.

#### 19.5.5.3 Critical evaluation summary

As I have stressed throughout this essay, the aesthetic processing theory has made contributions to HCI research and, I would argue, aesthetics research more generally. Its analysis of visceral perception and evaluation coupled with its ability to demonstrate subtle causal relations among diverse factors at the very heart of aesthetic perception are peerless as far as I know in any discipline. As Tractinsky notes, there is still much to do here, and as someone committed to aesthetic interaction and aesthetic life in general, I enthusiastically support the continuation of that agenda.

At the same time, as my list of weaknesses hopefully showed, there are major gaps the aesthetic processing theory's coverage of the whole domain of aesthetics, and these matter. Its reductionism, struggles with ethics and social context, lack of medium specificity, and problems with subjective experts and normative criteria are not merely incidental gaps that just haven't been filled in yet, but rather reflect intrinsic confounds in the theory itself. The existence of gaps and confounds is hardly a reason to reject a theory, since I can't think of a theory that doesn't have both.

But it *is* a good reason to be more epistemologically open-minded than Tractinsky portrays himself to be and to acknowledge as fellow travelers researchers who work on alternative formulations of aesthetics. Specifically, this encyclopedia chapter aiming to survey the whole field of visual aesthetic for HCI should acknowledge the history and philosophy of aesthetics as pursued throughout the humanities for millennia rather than replace them with vacuous dictionary entries, and also acknowledge work in HCI that builds on these traditions. The inclusion of such contributions would also enrich the prescriptions for the future of aesthetic research in HCI.

### 19.5.6 Aesthetics, according to the rest of the world

It has been a professional mystery to me, since moving to HCI from my doctoral work in comparative literature and philosophy, why so much of the work on aesthetics in HCI and design is so emphatically cut off from the rest of the aesthetic world. Indeed, reading this research, one might not even know that there is a massive domain of inquiry into aesthetics beyond aesthetic processing and other experimental traditions. Though I repeatedly stress that I find value in aesthetic processing, it is also worth pointing out to readers that aesthetic processing occupies the marginal position academically. If one consults the Wikipedia entry on aesthetics, or searches on the term “aesthetics” at Amazon or Google Scholar, what I am saying will become abundantly clear. Less than 10% of the Wikipedia article could even remotely be considered along the lines of what Tractinsky describes. Oxford’s 4-volume Encyclopedia of Aesthetics was edited by a philosopher (Michael Kelly) and is overwhelmingly not about aesthetic processing theory. Dozens of similar high profile examples in between these two extremes can be found.

I speculate that this self-imposed exile from millennia of interdisciplinary aesthetic thinking reflects a scientific habit that emerged in the Renaissance and came into its own in the Enlightenment and again in the Logical Postivism of the 1920s through 50s, which seeks to reject tradition as dogmatic and confused and seek instead to start anew, using rigorous science and empirically discovered facts to re-investigate phenomena that traditional culture muddled with its dogmas and pet theories. Yet the presence of traditional aesthetic concepts, categories, and systems of relations in aesthetic processing theory reveals the problem. If one uses vocabulary like “classical,” “expressive,” “aesthetic,” “experience,” and “judgment,” one is always already operating from inside the very tradition that is being rejected. And seeking to remove this vocabulary from aesthetic processing won't work—it can't be *aesthetic* processing unless some of this vocabulary is incorporated (and it always is).

So my argument instead is to accept the legitimacy of aesthetic processing but to end its self-imposed exile from the rest of aesthetics. (Humanists would benefit from a little empirical rigor as well—my argument cuts both ways.) What's needed, then, is to reintroduce the rest of the aesthetic world and juxtapose it to Tractinsky's essay, so that a fuller picture begins to emerge.

### 19.5.6.1 The aesthetic disciplines

As Tractinsky correctly argues, aesthetics is fundamental to human life and wellness. Not surprisingly, something so important to being human is going to get a lot of attention in human thought. One way to introduce it, then, is to focus primarily on contemporary aesthetic thought and the disciplines in which it unfolds. The following brief introduction is meant to sketch out what some of these disciplines are and the sorts of issues that people within them try to deal with. I also include a handful of introductory readings as starting points for those interested.

#### 19.5.6.1.1 Analytic philosophical aesthetics

Analytic aesthetics takes as its problem the careful evaluation of aesthetic systems of thought or dogmas. By way of self-disclosure, the analytic tradition has been most influential on my own understanding of aesthetics, and my approach to critiquing Tractinsky's essay is inspired by it. An analytic approach is typically strong at evaluating arguments about aesthetics, frequently (and frustratingly) demonstrating the irrationality of both everyday and sophisticated aesthetic positions. Nearly all of the core concepts of aesthetics (e.g., expression, pleasure, beauty, artworld, realism, experience, style, emotion, form, metaphor, representation, creativity, fiction) and aesthetic mediums (dance, poetry, film, theatre, painting, sculpture, etc.) are analyzed with logical rigor and definitional clarity in this discipline; it remains a mystery to me why analytic aesthetics isn't more influential in HCI than it currently is. Introductory works include Dickie's *Introduction to Aesthetics: An Analytic Approach* (1997), Eldridge's *An Introduction to the Philosophy of Art* (2003), and Levinson's *Oxford Handbook of Aesthetics* (2003), and for HCI readers I also recommend the first third of Carroll's *Beyond Aesthetics* (2001), because it speaks directly to major aesthetic issues in HCI today (including and especially formalist theories like Tractinsky's and experience theories like McCarthy & Wright's).

#### 19.5.6.1.2 Art history and theory

Aesthetics has historically been linked to reasoning about art. The art history and theory tradition is exceptionally strong at close analyses of art works, their experienced effects, the conditions of their creation, and the historical, national, cultural, and social contexts of their production and use. Beyond the innumerable large, full-color textbook histories of movements and traditions, I also recommend Julian Bell's *What is Painting?* (1999), Gayford & Wright's *The Grove Book of Art Writing* (1998), and any of the volumes in MIT Press' *Documents of Contemporary Art* series edited by Blazwick to get a feel for how art historians and artists think about art and aesthetics.

#### 19.5.6.1.3 Film aesthetics

Studies of film are of interest to interaction designers for several reasons. Film was the "new media" of the twentieth century, and one can see in the development of film and the social and intellectual reactions to it over time parallels to digital and interactive media and reactions today. As a dynamic, visual, and screen-based medium, film is also arguably closer to digital interaction than other cultural forms (such as novels or paintings). Finally, because film was so spectacularly implicated in the horrors of the twentieth century (Nazism in particular), film theorists and critics have intermingled aesthetic and ethical considerations in insightful ways that have similarly deep implications for HCI. Good starting points for interaction designers include Murch's *In the Blink of an Eye* (2001, also cited by Hassenzahl in his Commentary to this chapter), Braudy and Cohen's comprehensive *Film Theory and Criticism* (2004), and Monaco's visually exemplified introduction to film theory, *How to Read a Film* (2000).

#### 19.5.6.1.4 Literary theory

Literary theory has almost become synonymous with postmodernism and so-called "Grand Theory" (e.g., Deconstructionism, Psychoanalysis, Feminism, Marxism) and their excesses. One of the strengths (and problems!) with this tradition was its development of what has been called a "hermeneutics of suspicion," that is, the development of an interpretive habit that rejects traditional notions of authorial intention and aesthetic pleasure, and replaces them with analyses that claim to expose the secret machinations of the selfish subconscious, the false consciousness-creating ideologies of capitalism, and/or the repressive effects of patriarchy. Less acknowledged is that a backlash against Grand Theory within literary studies began in the late 1980s, and careful/close readings of literary texts has been making something of a comeback. I recommend as starting points Barry's *Beginning Theory* (2002), Tyson's *Critical Theory Today* (2006), and *The Cambridge History of Literary Criticism*, volumes VIII (Seldon, 2005) and IX (Knellwolf and Norris, 2007) in particular.

#### 19.5.6.1.5 Visual cultural studies

Visual cultural studies offers a cultural studies take on "both "high" and "low" forms of visual culture, including painting, product design, fashion, comic books, and advertising. These approaches are often interdisciplinary in their mix of critical interpretation, historical analysis, and sociological analysis and tend to remain well grounded on the visual artifacts and sociohistorical data, avoiding some of the dizzying flight of postmodernist "speculation to the death" (Baudrillard's phrase) characteristic of work in some of the other fields. The best introduction to this work I have seen is Barnard's *Approaches to Understanding Visual Culture*

(2001); also interesting are design histories, such as Marcus' *Masters of Modern Design: A Critical Assessment* (2005).

### 19.5.6.2 Cultural aesthetics and the human quality of life agenda

As I have spent much of my adult life reading about aesthetics from the fields listed above, I have my own (perhaps idiosyncratic) sense of the achievements of aesthetic thinking. I point out from the outset that these frequently comes in the form of theory—the kind of thing that Tranctinsky wants to rule out. But these theories are not, at least in the hands of the stronger writers, muddled and speculative dogmas; rather they are new concepts or new systems of concepts that empower us to perceive the experiential and socio-cultural significances of cultural works in much more diverse, nuanced, and personally fulfilling ways. These theoretical innovations are legitimated in at least two ways I can think of. First, they have to empower us to see and feel the sociocultural significance and experiential meanings of a work more robustly than we can without them. Second, they have to withstand, at least partly, the often brutal scrutiny of analytic philosophy, as described above.

I will briefly sketch some of the issues and related concept systems that have been developed to help us think more deeply about them—and to cultivate our appreciation for the aesthetic.

Whereas aesthetic processing theory formulates art as a "stimulus" that causes a "response" in our eyes and affective apparatus that in turn causes behavioral dispositions, other aesthetic theories position aesthetic encounters as the primary means by which an intelligent person works through a "learning-like process" to understand in an authentic and personal way "what it is like to live in the distinctive way of someone else," helping overcome the barriers to "coping with others" and also to overcome the "impoverishment of the sensitivity on which moral competence often depends" (Miller, 1998, p50). Here, Miller links **aesthetic perceptiveness** (which, incidentally, is a concept at the very core of

Baumgarten's definition of the word "aesthetics" when he invented the term in the 18th century) to an **empathic and holistic comprehension** of the **distinctive style of another's life** to **moral competence**. I can't in this space to justice to Miller's full reasoning, but his account is incredibly insightful and yet also participates elegantly in my own intuitive understanding of aesthetic experience. When I read Murakami's *Wind-Up Bird Chronicle* or watch Claire Denis' *35 Shots of Rum*, I don't merely enjoy the carefully crafted prose or sumptuous cinematography: I also feel like I am learning, expanding my horizons, growing in some way as a person. In short, Miller's account helps clarify what I already know: that aesthetic experience can be good for me.

But it is not always good for me. I spoke earlier of the "hermeneutics of suspicion," the interpretative strategy that with a dollop of paranoia investigates how aesthetic responses can be **false pleasures**. By *false pleasure*, I refer to pleasures that are harmful to us in ways that we fail to perceive or understand while we are enjoying them. Visually seducing desperate and fearful citizens to seek the paternal embrace of Nazism is a false pleasure, offering an **ideological myth** that stimulates the very fear it promises to assuage in place of actually providing a socially just system of government that would accomplish such security. Feminist film theorist Laura Mulvey 1975 showed in a seminal analysis how, in the hands of nearly exclusively male directors, cinema's camera imposes on viewers a heterosexual **male gaze**, that the camera visually inspects and finds scopophilic pleasures in female actors' bodies in ways that it does not for male actors, having diverse consequences for male and female film viewers.

Both of the preceding paragraphs have linked aesthetic experience with personal growth and the emergence of an intelligent and moral identity or the perversion of them. The following quote is from Richard Shusterman, an aesthetic philosopher who has influenced a considerable amount of HCI research to the point that Shusterman has been invited (and accepted) to be a featured speaker at CHI 2012. In the quote, Shusterman introduces his aesthetic vision:

"my prime goals here are reconstructive [i.e., normative] rather than historical: (1) to revive Baumgarten's idea of aesthetics **as a life-improving cognitive discipline** that extends far beyond questions of beauty and fine arts and that involves **both theory and practical exercise**; (2) **to end the neglect of the body** that Baumgarten disastrously introduced into aesthetics (a neglect intensified by the great idealist tradition in nineteenth-century aesthetics; and (3) to propose an enlarged, somatically centered field, somaesthetics, that can contribute significantly to many crucial philosophical concerns, thus enabling philosophy to more successfully redeem its original role as an **art of living**."

*-- Shusterman 2000; pp266-7, emphasis added*

Similarly, in the words of artist Nicolas Bourriaud:

"we are quite happy to create *modus vivendi* that make possible fairer social relations, more dense ways of life, and multiple, fruitful combinations of existence.... art no longer tries to represent utopias; it is trying to construct concrete spaces"

*-- Bourriaud 2006; pp 166-7*

My purpose in citing these quotes is not to assert that this is better or more important that the research ambitions of aesthetic processing theory, but simply to stress that this sort of agenda (a) is legitimately aesthetic HCI (including visually aesthetic HCI), because its aesthetic credentials are unassailable and it has influenced diverse HCI researchers (e.g., Schiphorst, 2009, 2011; Ferreira & Höök, 2009; Bardzell, 2011 among others), and (b) also can contribute to the same goals that aesthetic processing theorists themselves advocate: aesthetic interactions.

Aesthetic philosophers have also developed an extensive vocabulary to investigate **artistic expression**, including the development and articulation of sophisticated and intensely personal **emotional insights** (e.g., Collingwood, 1938) and the nature of **creativity**. Expression is important for HCI and interaction design, because unlike novels or paintings, interactions are made out of user expressions as much as designer choices. The aesthetic processing theory's focus on visceral perception/response rather than expression, reductiveness rather than holism, and lack of medium-specificity make it difficult (though not necessarily impossible) for this theoretical approach to offer a strong account of self-expression. Jinwoo Kim, in his Commentary to this chapter, suggests that it would be good to explore YouTube creativity and the "social formation process of visual aesthetics," and he is right. And, indeed, there is already HCI research on this that, again, is not acknowledged anywhere else in this chapter: Bardzell, 2007; Blythe & Cairns, 2009; Blythe & Cairns, 2010; Luther & Bruckman, 2008. I submit that this research agenda is strengthened, not confused, by a more holistic use of theory.

#### 19.5.6.3 Cultural aesthetics in HCI

As I have already suggested, Tractinsky's reference list notwithstanding, aesthetic theories from the humanities have influenced HCI and interaction design in significant and worthwhile ways. I briefly introduce a number of them below in the hopes of offering Encyclopedia of Human-Computer Interaction readers a more comprehensive and balanced set of references than they would have gotten had my Commentary not existed. I cannot here comprehensively cover all of the relevant work, but I do at least want to introduce four major themes of non-aesthetic processing approaches to aesthetics in HCI, themes that in many cases include seminal work.

The first research theme includes **medium-specific theories of interactive aesthetics**. As noted earlier, aesthetic processing does not explicitly distinguish between visual interaction and any other visual medium; it also scopes out non-visual digital interactions. Jonas Löwgren in (Löwgren, 2006, 2009) constructs an analytic vocabulary specifically for interaction, focusing on the notion of a *interaction quality*, which is a holistically understood description that intentionally blurs the boundaries between artifact descriptions and phenomenologically felt reactions. Examples include pliability, rhythm, dramaturgical structure, and fluency. In Bolter & Gromala (2006), the authors explore the ways that interactive technologies foreground and transform our understandings of transparency and reflectivity. In their analysis, these aesthetic qualities are also linked to dominant epistemologies in computer science. Lim et al. (2007) develop a concept of an *interaction gestalt*, leveraging both aesthetic processing (including Tractinsky) and pragmatist aesthetics (including Dewey and Shusterman) to articulate a set of attributes of an interaction gestalt, which include connectivity, continuity, movement, orderliness, pace, and time-depth among others. Common to all three medium-specific theories of interactive aesthetics are holistic understandings, explorations of design qualities, and efforts to link together interaction attributes with experience and understanding as they are consciously present to us.

The second research theme involves **design and research methodologies** surrounding aesthetic interaction. As interaction designers have gone from improving the performance of text editors to designing everyday technologies that are increasingly expected to be technologically robust, usable, sustainable, aesthetic, and socially just, design becomes an infinitely more complex problem. Standing in for empirical data about everything is an expert ability to read culture and to situate designs in appropriate and appealing ways within it (Kuutti, 2009). Supporting this ability to critically "read" design is a rising interest in the professional practice of criticism, an interest first introduced in HCI by Bertselsen & Pold (2004), and developed in my own work on *interaction criticism* (Bardzell & Bardzell, 2008; Bardzell 2009; and Bardzell, 2011). Also drawing on the arts and critical theory, but leveraging it in a different direction, is *critical design*, as developed in Dunne (2006), Dunne & Raby (2001), and Gaver et al. (2004). Critical designers develop provocative designs to challenge users by staging dilemmas that "force a decision onto the user, revealing how limited choices are usually hard-wired into products for us" (Dunne & Raby, 2001, pp45-46). Yet another approach, inspired by early twentieth-century literary theory, is *defamiliarization,* which seeks to find ways to enable designers to see beyond their own everyday assumptions by defamiliarizing themselves with and from them (Bell, Blythe, & Sengers, 2005).

A third research theme involves **specific aesthetic design domains** that touch upon HCI. In an anthology, a number of prominent HCI researchers from different intellectual traditions contributed a notion of *funology* as a new normative goal for HCI besides usability (Blythe et al., 2003). Another rising domain of aesthetic HCI interest is research on *craft and DIY*, which gets at a number of aesthetic issues, including creativity, the pleasures of things well made and making things well, and the relations among our pastimes, our heritage, and ourselves (e.g., Goodman & Rosner, 2011; Buechley et al., 2009; Bardzell, Rosner, & Bardzell, 2012). Responding to the explosion of non-WIMP-based user interfaces has been a rising interest in embodied interaction, and much of this work has also had an aesthetic dimension. Bardzell & Bardzell (2011) studied the designers and design processes of *digitally enabled sex toys* to understand designing for the confluence of sensual pleasure, intimate experience, social activism, and consumer electronics. HCI researchers have connected the *performing arts*, dance in particular, with interaction design research practices (Schiphorst, 2011) and user experience design and aesthetic response research (Latulipe, Carroll, & Lottridge, 2011).

Figure 19.4: High-end designer vibrators, such as the Better Than Chocolate Music Edition are aesthetic consumer electronics.



The final, and arguably most important, of all the themes I briefly sketch here is **aesthetic experience**. Aesthetic experience is a major topic in nearly all domains that identify themselves as "aesthetic," from recent analytic philosophy to the ancient Greeks, and from cognitive science to postmodern literary theory. Because it has been so tightly linked with user experience (UX) design—which remains the most common job title of my program's graduates—this one is of signal importance. The seminal work in this area is McCarthy & Wright's *Technology as Experience*, a book that constructs a theory of technology-mediated experience by combining American pragmatist philosopher John Dewey's theories of aesthetic experience in his seminal book *Art as Experience* with Russian literary theorist Mikhael Bakhtin's concept of dialogism. Their resulting theory has had extraordinary impact in our field not because (as Hassenzahl, Diefenbach, & Göritz complain) it seeks to tell us what the content of an experience is, but because the book has raised the imaginative perceptiveness, the insight, and the interpretative sensibility of thousands of UX designers (including all of my students)—a true achievement within the aesthetic HCI agenda. Others have used similar theory. In her recent *My Life as Night Elf Priest*, cultural anthropologist and HCI luminary Bonnie Nardi (2010) simultaneously constructs a theory of aesthetic experience and interpretively analyzes World of Warcraft play as an aesthetic experience. Her work helps interaction designers understand more analytically why World of Warcraft has been the smash hit that it has been, even as she contributes to the theory of aesthetic interaction experience. Finally, I want to point to Boehner, Sengers, & Warner's (2008) "Interfaces with the ineffable: Meeting aesthetic experience on its own terms," which among other contributions systematically explores the relationships between cognitive science and critical aesthetic conceptual systems. Though they position themselves as partisans on the critical side of that divide, and construct the two theoretical orientations as divided, nonetheless they can also be read subversively to explore opportunities to bridge that divide—which is increasingly what I think we should be doing.

### 19.5.7 A constructive conclusion

Given the unmistakeably increasing role of interaction in our everyday lives, mediating virtually every aspect of life, from work to the bedroom, interaction design simply must be aesthetic, just as our buildings must not only keep out rain but also be beautiful places to inhabit and our clothing must not only keep us warm but also help us express who we are or want to be. HCI is not mainly about high-performing text editors and aircraft controller interfaces any more. Whoever is seeking ways to make interaction more aesthetic I consider a fellow traveler.

We can't have tea for two without the two or the tea.

### 19.5.8 Acknowledgements

I'd like to acknowledge several scholars for helping me formulate the thoughts I have developed in this Commentary. I'd like to begin by thanking Noam Tractinsky, whose article prompted the strong feelings that all but compelled me to work out the ideas in this essay. While this article took on, at times, a critical tone, scholarly dialogue is the means by which knowledge advances are made, and my own thinking is clearer and (in my view) more beautiful than it would have been without his work on aesthetics in HCI. Methodologically, beyond acknowledging analytic philosophy in general, I'd like to reference Gilbert Cockton's careful critique of Gould & Lewis (2008) as an intellectual and rhetorical model for my analysis here. I'd also like to acknowledge Boehner et al.'s (2005) critique of the information processing model of affect in this, as that was the paper that introduced me to HCI's own "hermeneutics of suspicion" regarding the hidden operations of information processing theory in our research. Finally, I'd like to thank Mads Soegaard for convincing me to turn an emailed rant into a constructive commentary—and for his patience while I put it together.

### 19.5.9 References

- Bardzell, J. Creativity in amateur multimedia: popular culture, critical theory, and HCI. *Human Technology 3 (1)* (2007), 12-33.
- Bardzell, J. Interaction criticism and aesthetics. *Proc. of CHI'09*. ACM Press (2009), 2357-2366.
- Bardzell, J. Interaction criticism: An introduction to the practice. *Interacting with Computers 23* (6), 2011, 604–621.
- Bardzell, J. & Bardzell, S. Interaction criticism: a proposal and framework for a new discipline of HCI. In *CHI '08 Extended Abstracts*. ACM, New York, NY, 2463-2472.

- Bardzell, J., Bardzell, S. "Pleasure is your birthright": Digitally enabled designer sex toys as a case of third- wave HCI. *Proc. of CHI'10*, ACM Press (2011), 257- 266.
- Bardzell, S., Rosner, D., & Bardzell, J. Crafting Quality in Design: Integrity, Creativity, and Public Sensibility. *Proc. of DIS'2012*, ACM Press (2012, accepted).
- Barnard, M. (2001). *Approaches to Understanding Visual Culture*. New York: Palgrave Macmillan.
- Barry, P. Beginning Theory: An Introduction to Literary and Cultural Theory. Manchester, UK: University of Manchester Press, 2002.
- Bell, G., Blythe, M., & Sengers, P. Making by Making Strange: Defamiliarization and the Design of Domestic Technologies. *ACM Transactions on Computer-Human Interaction, Vol. 12,* No. 2, June 2005, 149–173.
- Bertelsen, Olav W. & Pold, S. Criticism as an approach to interface aesthetics. In *Proceedings of the third Nordic conference on Human-computer interaction* (NordiCHI '04). ACM, New York, NY, 23-32.
- Blazwick, I. Series Editor. *Documents of Contemporary Art*. Whitechapel Gallery and The MIT Press.
- Blythe, M., Overbeeke, K., Monk, A., & Wright, P. Funology: From Usability to Enjoyment. Dordrecht: Kluwer Academic Publishers, 2003.
- Blythe, Mark & Cairns, P. 2009. Critical methods and user generated content: the iPhone on YouTube. In *Proceedings of the 27th international conference on Human factors in computing systems* (CHI '09). ACM, New York, NY.
- Blythe, M. & Cairns, P. 2010. Tenori-on stage: YouTube as performance space. In *Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries* (NordiCHI '10). ACM, New York, NY, USA, 72-81.

- Boehner, K., DePaula, R., Dourish, P., and Sengers, P. Affect: From information to interaction. *AARHUS'05*, ACM Press (2005), 59-67.
- Boehner, K., Sengers, P., & Warner, S. Interfaces with the ineffable: Meeting aesthetic experience on its own terms. *ACM Transactions on Computer-Human Interaction* 15, 3, Article 12 (December 2008), 29 pages.
- Bolter, J. & Gromala, D. Transparency and reflectivity: Digital art and the aesthetic of interface design. In Fishwick, ed., *Aesthetic Computing*. Cambridge: The MIT Press, 2006, 369-82.
- Bourriaud, N. Relational aesthetics. In Bishop, C., ed., *Participation*. Cambridge, MA: The MIT Press, 2006, 160-171.
- Braudy, L. & Cohen, M. *Film Theory and Criticism*. Sixth edition. Oxford: Oxford University Press, 2004.
- Buchanan, R. Rhetoric, humanism, and design. Chicago: The University of Chicago Press, 1995, 23-66.
- Buechley, L., Rosner, D., Paulos, E., & Williams, A. 2009. DIY for CHI: methods, communities, and values of reuse and customization. In *Proceedings of the 27th international conference extended abstracts on Human factors in computing systems* (CHI EA '09). ACM, New York, NY, USA, 4823-4826.
- Carroll, N. *Beyond Aesthetics: Philosophical Essays*. Cambridge: Cambridge University Press, 2004.
- Collingwood, R. G. *The Principles of* Art. Oxford: Oxford University Press, 1938.
- Cockton, G. 2008. Revisiting usability's three key principles. In *CHI '08 extended abstracts on Human factors in computing systems* (CHI EA '08). ACM, New York, NY, USA, 2473-2484.
- Cross, N. *Designerly Ways of Knowing*. Basel: Birkhauser, 2007.

- Devereaux, M. Beauty and evil: The case of Leni Reifenstahl's *Triumph of the Will*. In Levinson, J., ed., *Aesthetics and Ethics: Essays at the Intersection*. Cambridge, UK: Cambridge University Press, 1998, 227-256.
- Dickie, G. *Introduction to Aesthetics: An Analytic Approach*. Oxford: Oxford University Press, 1997.
- Dunne, A. (2006). *Hertzian Tales: Electronic Products, Aesthetic Experience, and Critical Design*. Cambridge, MA: MIT Press.
- Dunne, A., and Raby. F. *Design Noir:The Secret Life of Electronic Objects*. Birkhäuser, Basel, Switzerland, 2001.
- Eldridge, R., 2003. *An Introduction to the Philosophy of Art*. Cambridge University Press, Cambridge, UK.
- Ferreira & Höök Body Orientations Around Mobiles CHI 2009
- Gaver, W., Bowers, J., Boucher, A., Pennington, S., Schmidt, A., Steed, A., Villars, N., Walker, B. The drift table: designing for ludic engagement. *Extended Abstracts of CHI*, ACM Press (2004).
- Gayford, M., & Wright, K. (eds). The Grove Book of Art Writing: Brilliant Writing on Art from Pliny the Elder to Damien Hirst. New York: Grove Press, 1998.
- Goodman, E. and D. K. Rosner. (2011) From Garments to Gardens: Negotiating Material Relationships Online and 'By Hand.' In *Proc. of CHI'11*.
- Greenberg, S. & Buxton, B. (2008). Usability considered harmful (some of the time). *Proc of CHI* ACM.
- Hassenzahl, M., Diefenbach, & Göritz, A. Needs, affect, and interactive products – Facets of user experience. *Interacting with Computers 22* (2010) 353–362

- Knellwolf, C. & Norris, C. (eds). *The Cambridge History of Literary Criticism: Volume IX Twentieth-Century Historical, Philosophical and Psychological Perspectives*, Cambridge: Cambridge University Press, 2001.
- Kuutti, K. HCI and design: uncomfortable bedfellows? In Binder, Lowgren & Malmborg (eds.) *(Re)searching the Digital Bauhaus*. London: Springer, 2009, 43-59.
- Latulipe, Celine, Carroll, E., A., & Lottridge, D. 2011. Love, hate, arousal and engagement: exploring audience responses to performing arts. In *Proceedings of the 2011 annual conference on Human factors in computing systems* (CHI '11). ACM, New York, NY, USA, 1845-1854.
- Levinson, J. *The Oxford Handbook of Aesthetics*. Oxford: Oxford University Press: 2003.
- Lim, Youn-kyung, Stolterman, E., Jung, H-K., & Donaldson, J. 2007. Interaction gestalt and the design of aesthetic interactions. In *Proceedings of the 2007 conference on Designing pleasurable products and interfaces* (DPPI '07). ACM, New York, NY, USA, 239-254.
- Löwgren, J. Articulating the use qualities of digital designs. In Fishwick, ed., *Aesthetic Computing*. Cambridge: The MIT Press, 2006, 383-404.
- Löwgren, J. (2009). Towards an articulation of interaction aesthetics. *The New Review of Hypermedia and Multimedia 15*(2), 129-146.
- Löwgren, J. & Stolterman, E. 2004 *Thoughtful Interaction Design*. MIT Press.
- Luther, Kurt and Bruckman, A. 2008. Leadership in online creative collaboration. In *Proceedings of the 2008 ACM conference*

*on Computer supported cooperative work* (CSCW '08). ACM, New York, NY, USA, 343-352.

- Marcus, G.H. *Masters of Modern Design: A Critical* Assessment. New York: Monacelli Press, 2005.
- McCarthy, J. and Wright, P. *Technology as Experience*. The MIT Press, MA, USA, 2004.
- Miller, R. Three versions of objectivity: Aesthetic, moral, and scientific. In Levinson, J., ed., *Aesthetics and Ethics: Essays at the Intersection*. Cambridge, UK: Cambridge University Press, 1998, 26-58.
- Monaco, J. *How to Read a Film: Movies, Media, Multimedia*. Third edition. Oxford: Oxford University Press, 2000.
- Mulvey, L. Visual pleasure and narrative cinema. *Screen* 16(3), 1975.
- Murch, W. *In the Blink of an Eye: A Perspective on Film Editing*. Second edition. Los Angeles: Silman-James Press, 2001.
- Nardi, B. *My Life as a Night Elf Priest: An Anthropological Account of World of Warcraft*. University of Michigan Press, Ann Arbor, 2010.
- Rogers, Y. (2006) Moving on from Weiser's vision of calm computing: Engaging ubicomp experiences. Ubicomp 2006, LNCS 4206, pp 404-421
- Schiphorst. T. 2011. Self-evidence: applying somatic connoisseurship to experience design. In *Proceedings of the 2011 annual conference extended abstracts on Human factors in computing systems* (CHI EA '11). ACM, New York, NY, USA, 145-160.
- Schiphorst. Thecla 2009. soft(n): toward a somaesthetics of touch. In *Proceedings of the 27th international conference extended abstracts on Human factors in computing systems* (CHI EA '09). ACM, New York, NY, USA, 2427-2438.

- Schön, D. A. *The reflective practitioner: How professionals think in action*. Basic books, 1983.
- Seldon, R. *The Cambridge History of Literary Criticism: Volume VIII From Formalism to Poststructuralism*. Cambridge: Cambridge University Press, 2005.
- Shusterman, R. *Pragmatist Aesthetics: Living Beauty, Rethinking Art*. 2000. Rowman & Littlefield.
- Tyson, L. *Critical Theory Today: A User-Friendly Guide*. Second edition. New York: Routledge University Press, 2006.


## 19.6 COMMENTARY BY GITTE LINDGAARD

How to cite this commentary in your report

### Gitte Lindgaard

*© Gitte Lindgaard*

Gitte Lindgaard, PhD, is a Distinguished Research Professor at Carleton University in Ottawa, Canada, and a Professor of Neuroaffective Design at Swinburne

University of Technology in Melbourne, Australia. Until recently, she was Director of the Human Oriented Technology Lab (HOTLab) holding the prestigious Canadian Natural Science & Engineering Research Council's NSERC/Cognos Senior In...

Gitte Lindgaard
Gitte Lindgaard is a member of The Interaction Design Foundation

### 19.6.1 Context, processes, and measurements of visual aesthetics in HCI: A commentary to Tractinsky's chapter on visual aesthetics

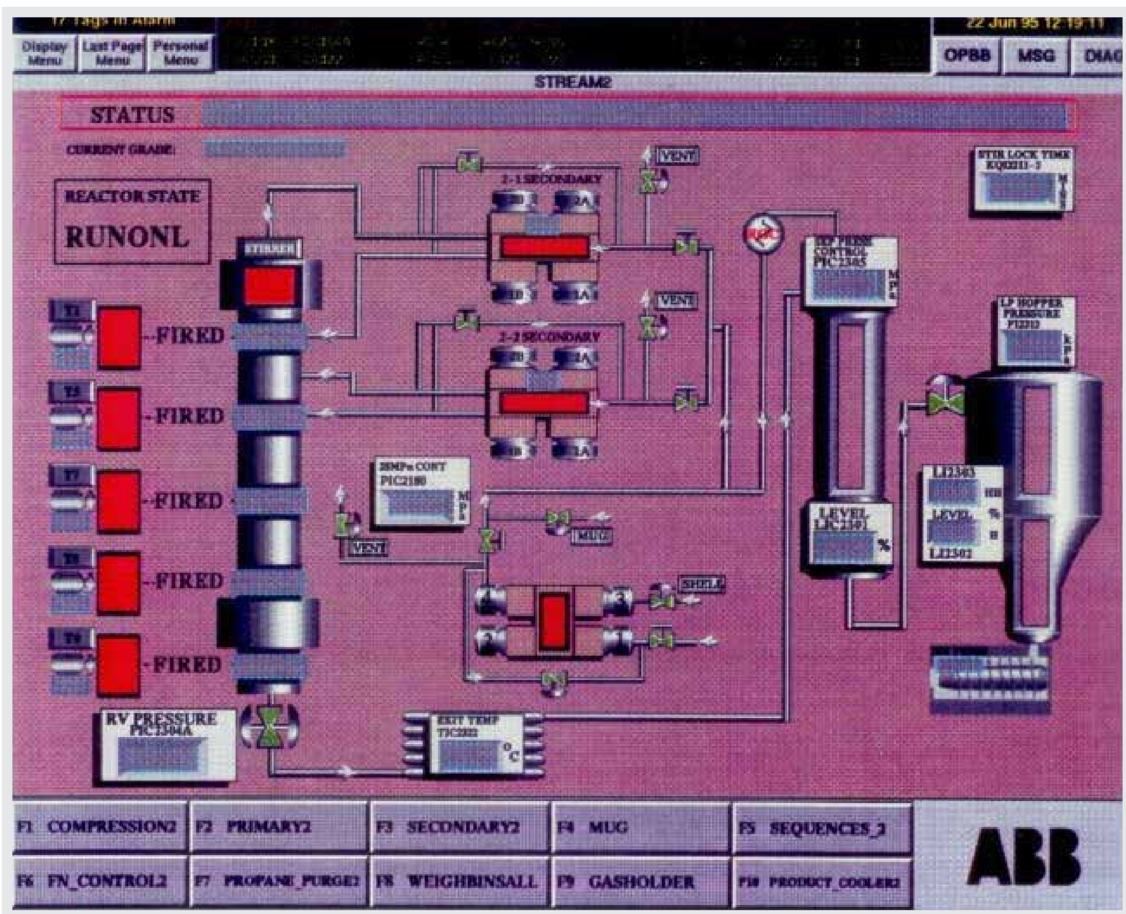
In his chapter, Tractinsky provides a thorough review of the aesthetics-related literature in the Human-Computer Interaction arena and beyond and it is a pleasure to read. It is especially nice to see just how far HCI research into visual aesthetics has come in 15 short years! Tractinsky reminds us of the origin of the concept of aesthetics and gives a very nice summary of relevant research from the perspectives of design and psychology as well as looking at practical issues of designed devices. In addition, Tractinsky also shares his views on where to go from here, outlining several strands of potential future research. I agree with most of Tractinsky offers in his essay, so I decided to extend some of the proposed directions. Specifically, I discuss the importance of context, people's expectations, and appropriateness with respect to visual aesthetics in an attempt to show that evaluation of aesthetics may occasionally be influenced by unrelated variables. The section following that discussion is a bold, tongue-in-cheek suggestion that it may be time for HCI researchers as well as product designers to consider the concepts of affect and cognition as an integrated whole, in addition to existing models and paradigms. I refer briefly to Barnard's Interacting Cognitive Subsystems (ICS) framework to underscore that the idea is not new. I provide research findings challenging the claim that the mere exposure effect is based entirely on affect. In the third section, I highlight some issues with one of Lavie and Tractinsky's (2004) aesthetics scales, the 'classical aesthetics' scale. Finally, I offer a conclusion.

### 19.6.2 The importance of context, expectations, and appropriateness of visual aesthetics

My thesis in this section is that context matters, even when we are interacting through a computer screen (Bødker, 1990), and even when our focus is on visual aesthetics. Computer games aiming to entertain and keep users engaged need vibrant colors, action-oriented settings, creative challenges and nifty surprises. Yet all of these attributes would be highly inappropriate for interactive technology designed, for example, to support the management of large-scale terrorist attacks involving mass casualties. Along similar lines, Web sites designed to facilitate the management of one's bank accounts should use graphics and color sparingly so as to look 'formal' and thus appear 'professional' (Lindgaard, Dudek & Fraser, 2012) and trustworthy (Kim & Moon, 1998). People don't go to the bank to be entertained or to hang out for extended periods of time. Yet, when looking for a gift for a special friend, the very same people who want banks to look formal expect lots of color and plenty of pictures displaying nicely presented products, perhaps even some playful animations. They enjoy spending time browsing an online gift shop that meets those expectations.

Over time, as experience with a particular website genre accumulates, our expectations of the look and contents of that genre develop into increasingly refined mental models (Johnson-Laird, 1983) or schemata (Bartlett, 1932), sometimes also referred to as look-up tables. These internal representations function as cognitive shortcuts by enabling us very quickly to determine how well a given exemplar of that genre meets our expectations. We tend to prefer the familiar, prototypical exemplars (Martindale, 1984; Winkielman, et al., 2006), mainly because they facilitate recognition and therefore demand a minimum of cognitive processing (Whitfield, 2000). To the extent that a particular website meets our expectation, we are likely to perceive it as an appropriate representative of its genre. To the extent that our expectations are not met, however, the site is likely to be deemed inappropriate even if it is well designed, very usable, and visually very appealing. In one of our recent experiments, we primed participants to expect to judge the visual appeal and appropriateness of a set of online banking sites or online gift shops even though they were all shown examples of both genres. The findings revealed that participants assigned to the gift shop condition rated visual appeal significantly higher than participants assigned to the banking condition (Lindgaard et al., 2012), and they were also significantly less tolerant of incongruent stimuli (Whittlesea & Williams, 2001). Mental models guiding expectations would thus seem to underlie the concept of appropriateness which, in turn, was shown to be capable of affecting perceptions of visual appeal. Although some HCI researchers have begun to investigate variables that may mediate perceptions and guide judgments of other variables (e.g. de Angeli et al., 2006; Moshagen & Thielsch, 2010; Hassenzahl & Monk, 2010; van Schaik & Ling, 2011), this research is in its infancy.

An interactive aesthetic experience is supposed to make us feel happy (Csikszentmihalyi, 1990), but as HCI researchers and designers we also need to understand what visual aesthetics means and what aesthetic experiences entail in a variety of situations. To date, nearly all visual-aesthetics related HCI research, including our own, has involved consumer goods or web sites. That is, research has focused on situations in which users decide themselves which products to buy and which websites to visit. Yet, it is equally relevant to consider aesthetics in the context of work where the choice of, and interaction with, technology is typically mandatory. In his research, Martindale (1990) found that meaningfulness was the most important predictor of preference. Meaningfulness may, however, on occasion lead to rejection of very appealing designs that, to the untrained eye, would be considered visually aesthetic and hence important for human well-being. For example, the images in Figure 1 below are borrowed from a high-pressure petrochemical plant-management system. The plant produces many types of plastic from purified, highly compressed gas injected under high pressure into reactor vessels operating at 200°+ C. The gas is mixed with chemical catalysts in a process that eventually outputs tiny plastic pellets forming the raw material for other products. Each of the four systems in the factory was represented by the very pretty, realistic 3-D graphical representation and by a different background screen color as shown in Figure 1. All four systems were accessible from the computer terminals, and the various parts of each system were directly accessible from those colorful front pages by clicking on the relevant component.



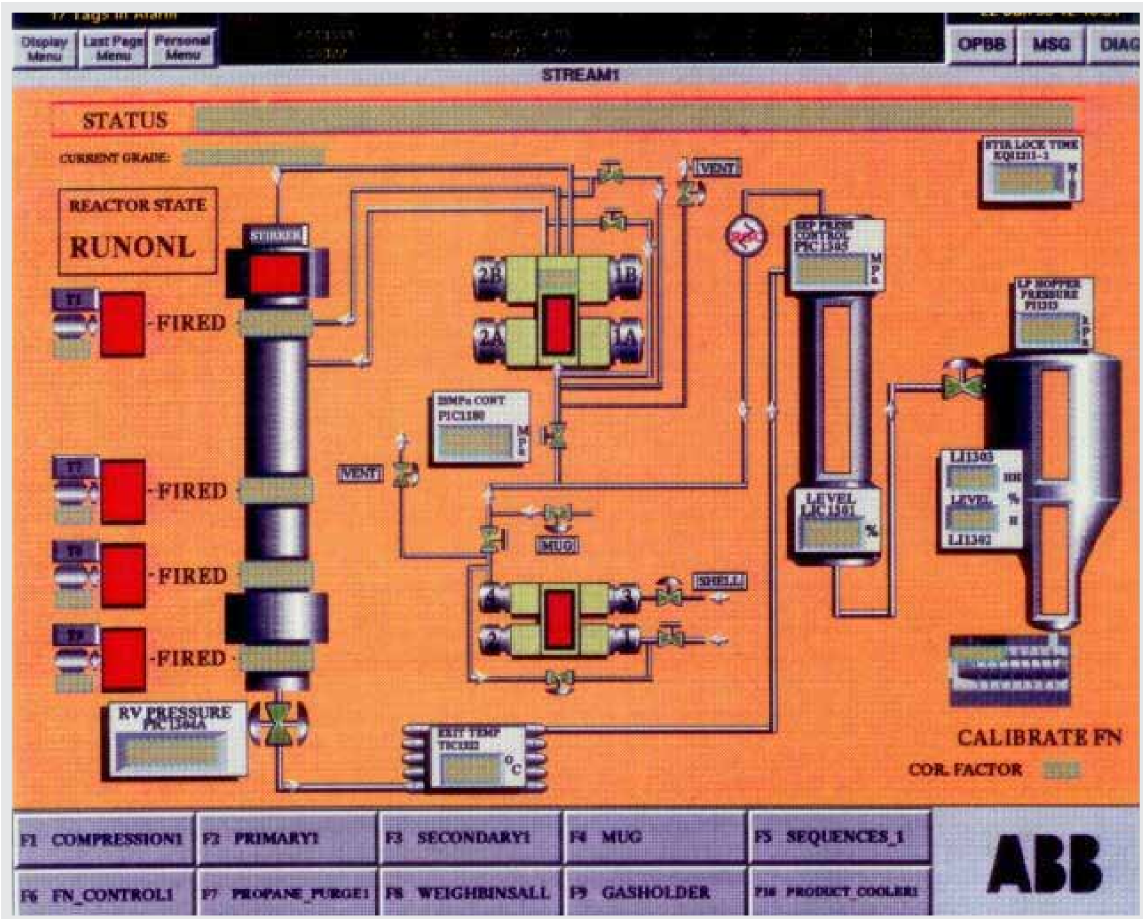





Figure 19.1 A-B: Screens representing two different systems in a high-pressure petrochemical factory. The first image has 5 pumps and 4 secondary compressors; the second image has 4 pumps and 2 secondary compressors (all with red borders).

Observations over several months of the highly experienced teams running the factory, however, showed that they did not use those screens to access the finer details of the systems. They noted only the number of pumps or the number of secondary compressors to ensure they were entering the intended system. When asked about the purpose of the different background colors, they maintained they 'hadn't noticed', and that to them, the background colors 'all looked pretty much the same'. To inspect components of a system, they used menus that relied on the terminology to which they were accustomed, or they reverted to the prototypical monochrome system diagram shown in Figure 2. The impressive graphic design efforts were, in other words, perceived to be unnecessary, indeed inappropriate, for that safety-critical environment.

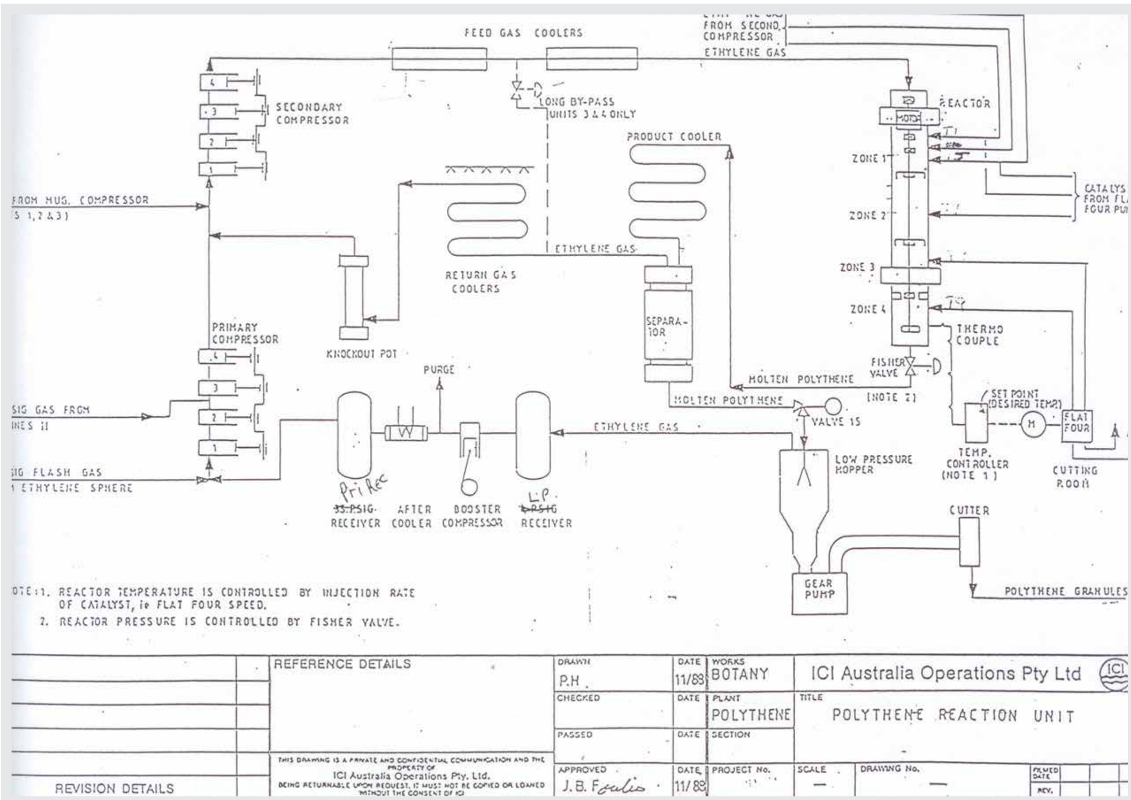



FIGURE 19.2: The paper diagram to which the experts reverted.



The above example highlights an important "(dis)connect" between designers and users, as Tractinsky so aptly puts it. Yet, in order visually to please a particular audience, images need not be 'pretty' in the conventional sense of everyone agreeing that they are 'good looking'. Images that may look very busy, even cluttered and thus not aesthetically pleasing to a lay audience, may be very pleasing and satisfying to work with for the target audience. The image in Figure 3 shows a screen that enables epidemiologists and infection control personnel effectively to monitor infectious disease outbreaks by tracing the people with whom affected patients may have been in contact since their exposure to the disease. This capability can thus also help to predict how the disease will spread unless preventative measures are taken such as isolating whole hospitals, even cities, in a timely fashion. To people whose work does not involve such issues, the screen may seem too bland and too busy; the target audience nevertheless finds it both visually appealing and useful.

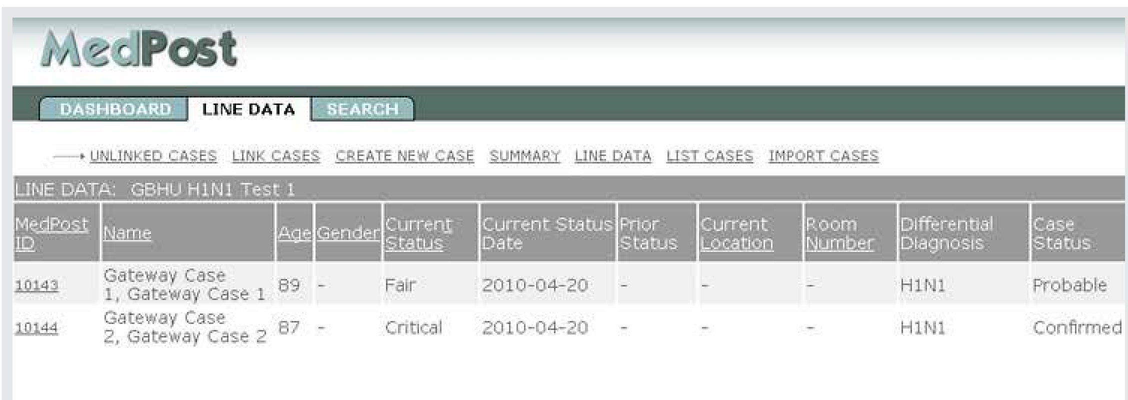



Figure 19.3: An epidemiologist's view of a screen allowing access to certain details about affected patients.



Both the above examples draw attention to the need to understand the meaning of visual aesthetics, its value to target users beyond the first impression, and the role it plays in different contexts. The issue is clearly more complex than merely deciding whether to impute or ignore visual aesthetics in the design of interactive technology as some researchers have speculated (Norman, 2004). To disentangle the roles of expectations and appropriateness in connection with visual aesthetics, we need longitudinal studies of ongoing interactive technology usage with self-chosen consumer products (Karapanos et al., 2009) as well as with mandatory systems in work places, targeting experts as well as new users.

### 19.6.3 Cognitive and affective processes

Hundreds of studies have confirmed the so-called mere exposure effect attributed to the work of Zajonc (1980; Bornstein 1989; 1992). It is found in experiments using a very brief stimulus exposure time, between one and 50 ms (Bornstein, 1989; 1992) in a variety of contexts including web pages (Lindgaard et al., 2006; 2011). The accumulated evidence suggests that it is based on affect and that it occurs in the absence of cognitive processes (Zajonc, 1980; 2001). According to Zajonc, "careful experiments have ruled out explanations of this phenomenon based on ease of recognition, and increased perceptual fluency, or subjective familiarity" (2001, p. 225). Zajonc further argues that, "if cognitive processes are not involved in a behavior… affective influences, which are necessarily less diverse than cognitive influences, will dominate the behavior, yielding a more homogeneous array of reactions" (2001, p. 227). Using a novel light-emitting diode (LED) tachistoscope, very recent research, however, has demonstrated that people are capable of recognizing and verbally identifying pictures of animals presented randomly for 1 or 10ms with mean levels of accuracy reaching approximately 90% (Thurgood et al., 2011). In one condition, the animal pictures were presented against a plain white background; in the other, they were shown in their natural environments. There was no difference in the number of animals correctly identified at 1 and 10ms exposure times in the plain condition, but more animals were correctly identified at 10 ms than at 1 ms exposure time in the natural-settings condition. The paradigm did not involve backward masking, the purpose of which is to cancel further processing of the target stimulus after its offset (Breitemeier & Ogmen, 2000; Verleger et al., 2004).

The proposed explanatory models of masking assume that the mask overrides the stimulus in the visual sensory buffer, replacing it with a representation of the mask. Rieger and his colleagues (2005) provided empirical support for this in a study in which they integrated psychophysical and physiological data and employed conditions with and without a mask. Using stimuli comprising complex images of natural scenes, their results showed that viewers had access to the stimulus beyond the target exposure time. Therefore, when no mask is used, it would appear that the iconic trace of the target stimulus remains in the visual buffer where it decays approximately one second after the stimulus offset (Averbach & Sperling, 1961; Kovacs et al. 1995; Sperling, 1960).

Due to the absence of masking, it is highly likely that Thurgood et al.'s (2011) results were affected, at least to some degree, by prolonged processing of the stimuli. However, contrary to previous findings involving the mere exposure effect, some cognition evidently did take place. As participants' responses were recorded manually, response times could unfortunately not be measured. Yet, Thurgood et al.'s research strongly suggests that we need to revisit our definitions of affect and cognition. If the two are as closely intertwined as these researchers' findings suggest, one may even speculate that the time has come to wean ourselves from the Cartesian dualism that has served science very well for several Centuries, but that demands us strictly to separate feeling from thinking. I believe it is time for us to start thinking about a more holistic view of human information processing that includes affect as well as cognition. Interestingly, Barnard's (1985) theoretical framework of Interacting Cognitive Subsystems (ICS) allows such smooth integration of affective and cognitive information (Barnard & Teasdale, 1991) that I have in mind. The central ICS concept is that different types of information are received, stored and processed by a set of nine functionally independent sub-systems whose function is to process sensory information, interpret it and prepare the organism to respond to events external to it (Humphrey, 1992). Because ICS is a framework rather than a theory, it makes no specific predictions about the exact representations used (Scott et al., 2001). 'Knowledge' is regarded as the consequence of several sub-systems functioning in a chain, whereby one passes the information to the next or to the outside world. A more complete explanation of ICS is given in Lindgaard and Whitfield (2004). I wholeheartedly agree with Tractinsky when he says that "the challenge is to identify and examine how various factors serve to alter or moderate the aesthetic process".

### 19.6.4 Measurements of visual aesthetics

Appraisals of visual aesthetics are typically obtained via rating scales (Hassenzahl, 2004; Lavie & Tractinsky, 2004; Moshagen & Thielsch, 2010, concurrent or retrospective verbal protocols (Ericsson & Simon, 1993; Taylor & Dionne, 2000), and/or psychophysiological measures (Jacobsen & Höfel, 2007a; 2007b; Tuch et al., 2009). Studies relying on rating scales feature most prominently in the HCI literature, and several of these have been found to correlate well with one another (see e.g. van Schaik & Ling, 2011; Moshagen & Thielsch, 2010), suggesting that they are tapping into the same concept. In his chapter, Tractinsky draws attention to the problematic issue of competing concepts that are not mutually exclusive, and which therefore causes confusion among researchers, students, and evaluators alike. The confusion concerns a conceptual overlap between Lavie and Tractinsky's (2004) 'classical' aesthetics scale and traditional usability.

Taking first a step back from visual aesthetics, 'Good design principles' have existed in the HCI literature since the 1980s (e.g. Smith & Mosier, 1985; Galitz, 1987; Ravden & Johnson, 1989), but most have their roots in human perception as discussed by the early gestalt psychologists (Koffka, 1915, cited in Köhler, 1967). Good design includes the so-called 'CRAP' Principles (Contrast, Repetition, Alignment, Proximity). Good contrast makes it easy for the eyes to distinguish between foreground and background. For example, the highly simplified white stick people in Figure 4 below stand out perfectly against the black background. Repetition refers to the use of a consistent visual system. For example, the same-size icons in Figure 4 denoting different kinds of sports all rely on a very simple visual language displaying drawings of one or two people shown in a frontal or a side view. The principle of alignment dictating a minimum number of alignment points is captured nicely in Tractinsky's Figure 19.3 of the two screens borrowed from Parush's (1998) study showing one very orderly screen that adheres to the alignment principle, and one very disorderly screen that does not. Likewise, proximity is also captured in the orderly screen in which items that belong together conceptually are placed together physically, with each group framed, and given a title that clearly distinguishes one the others. That is not the case in the disorderly screen in which individual items are more or less randomly placed.

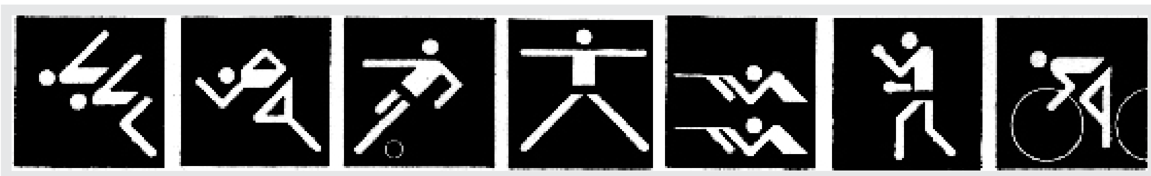


Figure 19.4: The principle of repetition via simplified drawings of people acting out a particular sport.

*Courtesy of Gitte Lindgaard. Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.*

These four basic design principles are largely adhered to in interactive computing systems regardless of whether an application is intended for serious or for more playful purposes, unless it specifically aims to confuse or surprise users, for example, in an interactive treasure hunt. The principles are also captured in four of the five the items in Lavie and Tractinsky's (2004) classical aesthetics scale. Thus, a pleasant, clear, and clean user interface design is well organized and orderly, much like Parush et al.'s (1098) good example. The role of symmetry, although recommended by some researchers (Sutcliffe, 2001; Bauerly & Liu, 2006), is a little unclear. For example, none of the icons in Figure 4 are horizontally or vertically symmetrical, but they are clear and clean, and they do reflect harmony. The final item in the classical aesthetics scale is 'aesthetics', which is somewhat curious in a scale intended to measure that very concept. In addition to being pleasant to look at, an orderly user interface design would also be easy to use and navigate. Those items, together with another item called 'clear design' feature in the additional scale intended to measure 'usability'.

It should be appreciated that Lavie and Tractinsky's two aesthetics scales were published nearly a decade ago and that they, together with Hassenzahl's (2004) scales, marked the first serious attempt in the HCI community to measure aesthetics such that concerns for visual aesthetics could be readily distinguished from traditional performance-based usability. The aesthetics-related scales have provided an excellent start allowing HCI researchers to delve more deeply into these complex concepts; they have served us well since their publication and have contributed to much fruitful research. Our next step now should be to conduct research aiming to resolve the unfortunate confusion about the conceptual overlaps between aesthetics, especially classical aesthetics, and usability.

### 19.6.5 Conclusion

Research into visual aesthetics has grown to become a very exciting, complex, and hence very challenging field in HCI. So many doors have been opened, many more topics are yet to be explored, and we have barely begun to identify some of the relevant concepts, let alone stray down some of those blind alleys that every new field of research inevitably will encounter. Thank you Noam, for reminding us of some of those directions we need to take, and thank you for summarizing the relevant literature for us.

### 19.6.6 References

- Averbach, E., & Sperling, G. 1961. Short term storage of information in vision. In C. Cherry (Ed.), *Information Theory*, (196-211). Butterworth, London.
- Barnard, P. J. (1985). Interacting cognitive subsystems: Modelling working memory, phenomena within architecture, in A. Miyake and P. Shah (eds), *Models of working memory: Mechanisms of active maintenance and executive control*, Cambridge University Press, Cambridge, 197—258.

- Barnard, P. J. & Teasdale, J. D. (1991), Interacting cognitive subsystems: a systematic approach to cognitive-affective interaction and change, *Cognition and Memory*, 5, 1—39.
- Bartlett, E.F. (1932). *Remembering*, Cambridge University Press, Cambridge, UK.
- Bauerly, M. & Liu, Y. (2006). Computational modeling and experimental investigation of effects of compositional elements on interface and design aesthetics, International Journal of Human-Computer Studies, 64 (8) pp. 670-682
- Bornstein, R.F. (1989). Exposure and affect: Overview and meta-analysis of research, 1968-1987, *Psychological Bulletin*, 106(2), 266-289.
- Bornstein, R.F. (1992). Subliminal Mere Exposure Effects, In R. F. Bornstein & T. S.Pittman (eds), *Perception Without Awareness: Cognitive, Clinical, and Social Perspectives*, The Guilford Press, New York, 191-210.
- Bødker, S. (1990). *Through the interface: A human activity approach to user interface design*, Lawrence Erlbaum Associates, Mahwah, NY.
- Breitemeier, B.G. & Ogmen, H. (2000). Recent models and findings in visual backward masking: a comparison, review, and update, *Perception and psychophysics*, 62 (8), 1572-1595.
- Csikszentmihalyi, Mihaly (1990). *Flow: The Psychology of Optimal Experience*. Harper & Row, New York, NY.
- DE ANGELI, A., SUTCLIFFE, A. AND HARTMANN, J. 2006. Interaction, usability and aeshtetics: What influences users' preferences?, *Proceedings DIS*, 271-280.

- Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data* (Rev. ed.). MIT Press, Cambridge, MA.
- HASSENZAHL, M, 2004. The interplay of beauty, goodness, and usability in interactive products, *Human-Computer interaction*, 19, 4, 319-349.
- Galitz, W. (1987).
- Hassenzahl, M. & Mohk, A. (2010). The inference of perceived usability from beauty, *Human-Computer Interaction*, 25(3), 235-260.
- Humphrey, N. K. 1992, A History of the Mind (London: Chatto and Windus).
- JACOBSON, T. AND HÖFEL, L. 2007a. Electrophysiological indices of processing aesthetics: Spontaneous or intentional processes, *International journal of psychophysiology*, 65, 20-31.
- JACOBSON, T. AND HÖFEL, L. 2007b.Electrophysiological indices of processing symmetry and aesthetics: A result of judgment categorization or judgment report?, *Journal of psychophysiology*, 21 (1), 9-21.
- Johnson-Laird, P. (1983). *Mental models*, Harvard University Press, Cambridge, MA.
- Köhler, W. (1967). Gestalt Psychology, Psychological Research, 1, XVIII-XXX
- KOVACS, G., VOGELS, R. AND ORBAN, G.A. 1995. Cortical correlate of backward masking, *Proceedings of the national academy of sciences U.S.A.*, 92, 5587-5591.
- Karapanos, E., Zimmerman, J., Forlizzi, J., & Martens, J. B. (2009). User experience over time: an initial framework *CHI2009* (pp. 729-738): ACM.

- Kim, J. & Moon, J.Y. (1998). Designing towards emotional usability in customer interface trustworthiness of cyber-banking system interfaces, *Interacting with Computers*, 10, 1-29.
- Klein, G. (1997). An overview of naturalistic decision making applications, in X. Zambok & G. Klein, Naturalistic decision making, Lawrence Erlbaum & Associates,
- Lavie, T. & Tractinsky, N. (2004). Assessing dimensions of perceived visual aesthetics of web sites, *International Journal of Human-Computer Studies*, 60, 269-298.
- Lindgaard, G. & Whitfield, A. (2004). Integrating aesthetics within an evolutionary and psychological framework, *Theoretical Issues in Ergonomics Science,* 5 (1), 73-90.
- Lindgaard, G., Fernandes, G., Dudek, C. & Brown, J. (2006). Attention web designers: You have 50 milliseconds to make a good first impression!, *Behaviour & Information Technology*. 25, 115-126 (awarded best paper for 2006, 25th. Anniversary of BIT).
- Lindgaard, G., Dudek, C., Sen, D., Sumegi, L. & Noonan, P. (2011). An exploration of relations between visual appeal, trustrworthiness and perceived usability of homepages*, TOpics in Human-Computer Interaction (TOCHI)*, 18(1), 1-30.
- Martindale, C., Moore, K. & Borkum, J. (1990). Aesthetic preference: Anomalous findings for Berlyne's psychobiological theory, *The American journal of psychology*, 103(1), 53-80.
- Lindgaard, G., dudek, C. & Fraser, L. (2012). Exploring the role of mental models on judgments of visual appeal and appropriateness of web pages, submitted to International Journal of Human-Computer Studies (February, 2012).

- Martindale, C. (1984). The pleasures of thought: A theory of cognitive hedonics, *Journal of mind and behavior*, 5, 49-80.
- Moshagen, M. & Thielsch, M.T. (2010). Facets of visual aesthetics, *International journal of human-computer studies*, 68,689-709.
- Norman, D.A. (2004). *Emotional design: Why we love (or hate) everyday things*, Basic Books, New York, NY.
- Ravden & Johnson
- RIEGER, J.W., BRAUN, C., BÜLTHOFF, H.H. AND GEGENFURTNER, K.R. 2005. The dynamics of visual pattern masking in natural scene processing: A magnetoencephalography study, *Journal of vision*, 5, 275-286.
- Roth, S.P., Schmutz, P., Pauwels, S.L., Bargas-Avila, J.A. & Opwis, K. (2010). Mental models for web objects: Where do users expect to find the most frequent objects in online shops, news portals, and company web pages?, *Interacting with Computers*, 22, 140-152.
- Scott, S. K., Barnard, P. J. and May, J. 2001, Specifying executive representations and processes in number generation tasks, The Quarterly Journal of Experimental Psychology, 544, 641—664.
- Smith & Mosier (1985)
- SPERLING, G. 1960. The information available in brief visual presentations. *Psychological Monographs: General and Applied*, 74(11), 1-30.
- SUTCLIFFE, A. 2001 Heuristic evaluation of website attractiveness and usability, *Lecture Notes in Computer Science (LNCS) 2220*, 183-198.
- Taylor, K.L. & Dionne, J.-P. (2000). Accessing Problem-Solving Strategy Knowledge: The Complementary Use of Concurrent Ver-

bal Protocols and Retrospective Debriefing, *Journal of educational psychology*, 92(3), 413-425.

- Thurgood, C., Whitfield, T. W. A., & Patterson, J. (2011). Towards a visual recognition threshold: New instrument shows humans identify animals with only 1 ms of visual exposure. *Vision Research, 51*(17), 1966-1971.
- Tuch, A.N., Bargas-Avila, J.A., Opwis, K. & Wilhelm, F.H. (2009). Visual complexity of websites: Effects on users' experience, physiology, performance, and memory, *International Journal of Human-Computer Studies,* 67, 703-715.
- van Schaik, P. and Ling, J. (2011). An integrated model of interaction experience for information retrieval in a Web-based encyclopaedia, *Interacting with Computers*, 23 (1), pp.18-32.
- VERLEGER, R., JASKOWSKI, P., AYDEMIR, A., VAN DE LUBBE, R.H.J. AND Groen, M. 2004. Qualitative differences between conscious and non-conscious processing? On inverse priming induced by masked arrows, *Journal of experimental psychology: General*, 133, 494-515.
- Whitfield, T. W. A. (2000). Beyond prototypicality: towards a Categorical-Motivation model of aesthetics, *Empirical Studies of the Arts*, 18, 1—11.
- Whittlesea, B.W.A. & Williams, L.D. (2001). The discrepancy-attribution hypothesis: II. Expectation, uncertainty, surprise, and feelings of familiarity, *Journal of experimental psychology: Learning, memory, and cognition*, 27(1), 14-33.
- Williams, R. (2008). *The non-designer's design book*, Peachpit Press, Berkeley, CA. 3rd edn.

- Winkielman, P., Halberstadt, J., Fazendeiro, T. & Gatty, S. (2006). Prototypes are attractive because they are easy on the mind, *Psychological science*, 17(9),799-806.
- Zajonc, R., 1980. Feeling and Thinking: Preferences need no Inferences, *American Psychologist*, *35* (2), 151-175.
- Zajonc, R.B. 2001. Mere exposure: A gateway to the subliminal, *Current directions in psychological science*, *10* (6), 224-228.


## 19.7 COMMENTARY BY MARC HASSENZAHL

How to cite this commentary in your report


### Marc Hassenzahl

Marc Hassenzahl is Professor at the Folkwang University in Essen and research manager at MediaCity, Åbo Akademi University, Vaasa, Finland. He is interested in the positive affective and motivational aspects of interactive technologies – in short: User Experience...

Marc Hassenzahl
Marc Hassenzahl is a member of The Interaction Design Foundation


### 19.7.1 Everything can be beautiful

“Beauty is an important ingredient of our daily lives. We admire and praise the beauty of nature, architecture, music, other people — an ugly colour or an awkward form easily repels us. Given its pervasiveness, the lack of research addressing aesthetics in Human-Computer Interaction (HCI) is striking”.

Not long ago, I started a book chapter on beauty and HCI with these words (Hassenzahl, 2008). And I believe both parts still to be true. Beauty still matters and HCI still keeps struggling with the concept. The alleged reasons for this are manifold. One can despise beauty because of its notorious elusiveness or fight about whether beauty can be reduced to some numbers on scales or not — as David Frohlich (2004) put it: "It didn't seem to me to be the kind of thing that could be measured so easily with a seven-point bipolar scale and a pencil". Not to mention the legions of philosophers who already devoted whole lives to understanding the transcendental nature of beauty.

Noam Tractinsky is not easily deterred by this. He boldly summarizes what we know about beauty in HCI, which seems to be substantial enough. We know, for example, something about the processes underlying judgments of beauty. When "looking at" an object the percept is emotionally processed. This leads to a positive or negative response — an involuntary, effortless and fast process. Attributed to the visual Gestalt of an object, the response becomes its beauty. We can more or less reflect and elaborate upon this initial response and we can even revise it.

But more importantly any judgment of beauty has consequences. Through its immediacy, beauty becomes *the* starting point for inferring other attributes, such as how practical or captivating an object is — even when actual hands-on experience is missing (van Schaik, Hassenzahl, & Ling, in press). "What is beautiful is ..." is a powerful process, and trying to understand when and how people infer quality through a network of beliefs and rules is exciting. An even more striking phenomenon is the ambivalent nature of beauty in the consumers' eye. Sarah Diefenbach calls it the "Beauty Dilemma" (e.g., Diefenbach & Hassenzahl, 2011). In fact, we all seem to know how much we enjoy beauty. According to Maslow (1954) beauty might even be a fundamental need (which I do not necessarily agree with) and Raymond Loewy — "the man who streamlined the sales curve" — endowed us with the insight that "between two products equal in price, function and quality, the one with the most attractive exterior will win." Nevertheless, Sarah finds a deeply ingrained suspicion towards beauty in products. Choosing a primarily beautiful over a primarily usable product is difficult, because it needs to be justified. We want beauty, but we are desperately looking for any "functional alibi" easing the load of justifying our desire. That is why *Apple* users insist that their gadgets are not only beautiful but also more usable. It is a proper justification for indulging in beauty. There are other envisioned consequences of beauty even that "attractive things work better", and knowing those seems important for any discipline concerned with making things. We cannot switch off peoples' perception and evaluation of the things in their environment, thus, we cannot *not* address beauty (or ugliness, respectively) when designing. We better know of the consequences of ignoring beauty.

Coming back to the beginning of my comment: Why are we struggling with beauty, given that we already know so much? Close-up beauty seems only half as elusive as it appeared at the outset and the many interesting and important consequences make research into beauty valuable. In a comment to one of my papers on beauty, Kees Overbeeke and Stephan Wensveen (2004) stated: "For product designers Hassenzahl's work is of interest [...] if it can be used in actual design work. How will it contribute to new product development?" Now substitute "Hassenzahl's work" with "research on aesthetics" and you see the problem. Knowing the processes of how to derive a judgment of beauty or the consequences once it was derived tells us nothing about how to make something beautiful. To learn more about how to make beautiful things, I consulted the "antecedents section" of Noam's paper, but it leaves me empty-handed. I trust in Paul Hekkert's (Hekkert, Snelders, & van Wieringen, 2003) advice to balance typicality and novelty — but what exactly *is* typicality and novelty then, doesn't this just only shift the problem? — and I mistrust the potential helpfulness of advice such as "colour use should be balanced and low saturation pastel colours should be used for backgrounds ..." (Sutcliffe, 2009). I hate pastels — most of the time.

"If we could only decipher the aesthetic code!" Noam exclaims — ironically maybe — but to me it reveals the basic problem. In fact, there is nothing to decipher, *we simply make beauty*. Given the swiftness of the judgmental processes, many academic quickly invoke innate mechanisms shaped by evolution as explanation for beauty. That's how the argument goes: We respond favourably to symmetry, because it signals health (i.e., reproductive success). And somehow we fail to recognize that a TV set or a car has only a weak relation to reproduction — we still like it better when it is symmetric — maybe. Evolutionary explanations are hard to rebut, but actually I don't think we need them. Let's think of the judgmental process underlying beauty as a short-hand. It is one of those magically fast, automatic System 1 processes that spare our lazy System 2 the deliberate thinking (Kahneman, 2011). Without such short-hands or "heuristics", we would be catatonic most of the time — locked into endless choice processes.

But even when we think of judgments of beauty as a short-hand, the crucial question remains: Why do we react, in a split-second, to one object positively, but negatively to the other? It's not exactly an original observation, but I suspect this to be first of all a matter of *familiarity* (e.g., "mere exposure", Zajonc, 1980). For car interiors, Carbon and Leder (2005), for example, showed that highly innovative designs were not judged to be beautiful at first. However, repeated unobtrusive exposure (over 30 minutes) quickly increased beauty. The other important aspect is *authority*. It is not an immediately perceivable inherent quality that distinguishes a design classic from any other object. It is the very fact that accepted authorities announce it to be a design classic— through exhibiting, reviewing, and giving away precious awards — which counts.

Without familiarity or authority guiding us through unfamiliar masterpieces we have a hard time to perceive beauty. A good example is the one of the street musician playing his violin for 43 minutes on a Friday morning at *L'Enfant Plaza*, Washington, without attracting much attention. Hardly anybody stopped; the youngish man collected $32. The musician, however, was Joshua Bell, the violin a $3.5 million handcrafted Stradivari from 1713, and the music masterpieces by Bach, Schubert, Ponce and Massenet. The same people that passed Bell in the metro without a second look may pay $100 for an admission ticket to listen to him in a concert hall.


Video 1: Joshua Bell, an American Grammy Award-winning violinist plays his $3,5 million Stradivari violin and earns $32 in 43 minutes.


Video 2: Joshua Bell playing his violin at the White House Evening of Classical Music on November 4, 2009.

In the Washington Post article about the metro experiment, Mark Leithauser, senior curator at the National Gallery, makes it clear: “Let’s say I took one of our more abstract masterpieces, say an Ellsworth Kelly, and removed it from its frame, marched it down the 52 steps that people walk up to get to the National Gallery, and brought it into a restaurant. It’s a $5 million painting. And it’s one of those restaurants where there are pieces of original art for sale, by some industrious kids from the Corcoran School, and I hang that Kelly on the wall with a price tag of $150. No one is going to notice it. An art curator might look up and say: ‘Hey, that looks a little like an Ellsworth Kelly. Please pass the salt.’” One may attribute this to context; I think it is about authority.

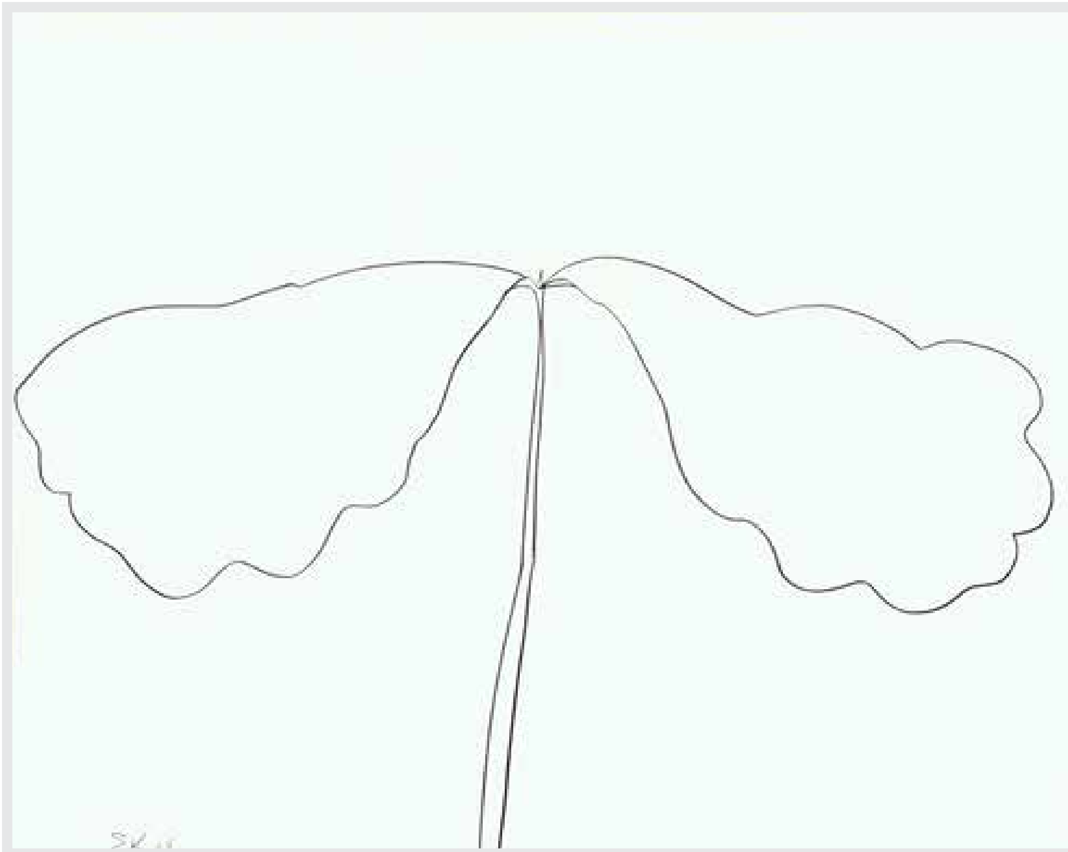


FIGURE 19.1: Ellsworth Kelly *Small Oak*, 1964, National Gallery, Washington, Not on View.



To make something beautiful is thus not about curves versus rectangles, saturation, hue, symmetry, proportions or any other hidden "aesthetic code." To make something beautiful is about deciding what to want, to make it, to expose people to it, and to claim with authority that this is beautiful. In this respect beauty is more or less constructed socially. For design this is freedom and burden at the same time. While we can make everything beautiful — even streamlined toasters — we become more and more aware of the responsibility this implies. It was us and not any evolutionary aesthetic code, who established the wasp waist, subjecting women to cracked and deformed ribs, weakened abdominal muscles, and deformed and dislocated internal organs. Was Rubens just depicting the beauty ideal of his time or was he actually *setting* it to voluptuous, stout, and luxuriant? Is it some hard-wired evolutionary preference or us, who decided to create a beauty ideal in cars that look as if they run on chummy pedestrians rather than on gasoline? Because everything can be beautiful, we need to think careful about what we make beautiful, how we set our ideals. This is the true challenge of beauty in HCI and any other design discipline.

## References


- Carbon, C.-C., & Leder, H. (2005). The Repeated Evaluation Technique (RET). A method to capture dynamic effects of innovativeness and attractiveness. *Applied Cognitive Psychology*, *19*(5), 587-601. doi:10.1002/acp.1098
- Diefenbach, S., & Hassenzahl, M. (2011). The Dilemma of the Hedonic - appreciated, but hard to justify. *Interacting with Computers*, *23*(5), 461-472. doi:10.1016/j.intcom.2011.07.002
- Frohlich, D. (2004). Beauty as a Design Prize. *Human-Computer Interaction*, *19*(4), 359-366. Taylor & Francis. doi:10.1207/s15327051hci1904_4
- Hassenzahl, M. (2008). Aesthetics in interactive products: Correlates and consequences of beauty. *Product experience* (pp. 287-303). Retrieved from http://issuu.com/hassenzahl/docs/beauty-exp-hassenzah-lit-fin
- Hekkert, P., Snelders, D., & van Wieringen, P. C. W. (2003). "Most advanced, yet acceptable": typicality and novelty as joint predictors of aesthetic preference in industrial design. *British journal of psychology (London, England : 1953)*, *94*(Pt 1), 111-24. doi:10.1348/000712603762842147

- Kahneman, D. (2011). *Thinking Fast & Slow* (p. 512). Allen Lane. Retrieved from http://www.amazon.com/Thinking-Fast-Slow-Daniel-Kahneman/dp/1846146062
- Maslow, A. H. (1954). *Motivation and Personality*. (A. H. Maslow, Ed.)*New York* (Vol. 19, p. 411). Harper. Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Motivation+and+Personality#0
- Overbeeke, K., & Wensveen, S. (2004). Beauty in Use. *Human-Computer Interaction*, *19*(4), 367-369. Taylor & Francis. doi:10.1207/s15327051hci1904_5
- Sutcliffe, A. (2009). Designing for User Engagement: Aesthetic and Attractive User Interfaces. *Synthesis Lectures on Human-Centered Informatics*, *2*(1), 1-55. doi:10.2200/S00210ED1V01Y200910HCI005
- Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist*, *35*(2), 151-175. doi:10.1037/0003-066X.35.2.151
- van Schaik, P., Hassenzahl, M., & Ling, J. (n.d.). Modeling user-experience from an inference perspective. *Transactions of Computer-Human Interaction*.

## 19.8 COMMENTARY BY ANTONELLA DE ANGELI

How to cite this commentary in your report

### Antonella De Angeli

© *Antonella De Angeli*

I am an Associate Professor of Human-Computer Interaction at the Department of Information Engineering and Computer Science of the University of Trento in Italy. My research addresses cognitive, social and cultural aspects of information technologies with an emphasis on the application of this knowledge to interaction design....

Antonella De Angeli
Antonella De Angeli is a member of The Interaction Design Foundation

Antonella's commentary is coming very soon - please check back in a few days!

## 19.9 COMMENTARY BY DIANNE CYR

How to cite this commentary in your report

## Dianne Cyr

© *Dianne Cyr*

Dianne Cyr is Professor at Simon Fraser University. My background is varied and interdisciplinary. Both my Bachelor and Masters degrees are in Psychology, and I worked in clinical psychology for the better part of a decade before returning to university to embark on doctoral studies. The earlier training held me in good stead for my Ph.D. research which was focused on the linkage of s...

Dianne Cyr
Dianne Cyr is a member of The Interaction Design Foundation

Noam Tractinsky uses aesthetics to refer to an artistically or pleasing appearance or effect. There is a focus on the visual and the attributes that encroach on our hedonic or affective sensibilities. As an art collector for many years, a new abstract acquisition will move me in certain ways - based on the colors, balance and overall flow of the work. But as with visual aesthetics or visual design in HCI, it is often difficult to explain why. While it may no longer be heretical to consider design or visual aesthetics as central user elements in HCI, in-depth understanding of these areas remains elusive — especially from a theoretical perspective.

Tractinsky's contributions to the field of design are substantial, and he has provided an impetus through his research to better understand the design perspective and what is pleasing to the user. His article utilizes three lenses— the design perspective, the psychological perspective, and the practical perspective — all of which have implications for the ongoing development of theory that informs practice. Building from this work, I propose there are three areas that merit additional attention by HCI researchers and practitioners as we move forward in the field, and it is these areas which I address below. That is: (1) the need to theoretically ground the principles of visual aesthetics in experimentally driven research; (2) the expansion of methodologies to gain new perspectives on visual aesthetics and design, and (3) new directions for research that emphasize individual and cultural factors. These are all topics mentioned in this chapter.

### 19.9.1 Developing Theoretical Models for Design Aesthetics

In 2007 Shirley Gregor and David Jones wrote an article in the *Journal of the Association for Information Systems* titled "The Anatomy of a Design Theory". The crux of the paper is that insufficient attention has been focused on the specification of design theory in terms of identifying purpose and scope; constructs to be tested; principles of form and function that define the structure, organization, and functioning of the design product or design method; and principles of implementation, among others. While great strides have been made in these areas in recent years — there is room for improvement to carefully identify and test principles of design aesthetics, and the subsequent impact on the user. Perhaps this is, as Tractinsky points out, because aesthetics and other design principles are intertwined and the specifics of design as tested are not sufficiently delineated. As is also mentioned in this chapter, terms such as "aesthetics" or "beauty" are ill defined, originating in diverse disciplines in which different meanings of the terms prevail. The psychological perspective further exacerbates the development of theory since it is difficult to determine precisely why a user responds to design elements. Along these lines, what exactly is pleasure? How is it different from enjoyment or satisfaction? Although some studies are aiming to disentangle these various constructs as reactions to design, in the realm of affect and emotion there is much work yet to be done. Developing or adopting theoretical frameworks that underpin such emotive or affective responses is essential to the development of design theory that can be tested over time in diverse contexts. For instance, in our paper in which we examined human images in website design (Cyr et al. 2009) images were examined as they contributed to user's perceived social presence of the website. The theory of visual rhetoric was used as the theoretical context for user experience and provided understanding for why emotive responses occur. In other work it is likewise important to carefully outline theory, apply it to user experience, and to build new understandings that add rigor to the discipline.

### 19.9.2 Methodologies through which to Interpret Visual Design

Related to the preceding, it is generally acknowledged that methodologies are best used when they are suited to the research problem, and are aimed to elicit depth and precision. In this regard, Tractinsky aptly profiles various methods for how user responses to visual aesthetics or visual design are gauged. He notes that typically visual design is measured using surveys with single or multiple item scales. However as measurement techniques expand, there is opportunity to delve into new methods that more deeply and comprehensively attend to what users are experiencing. More specifically, Tractinsky refers to the study using fMRI for testing reactions to product packaging (i.e. Reimann et al., 2010), and there is merit to pursue these alternative methodologies as they inform the HCI and design communities. For instance, in our work we examined human images in website design (Cyr et al. 2009) as well as user reactions to the use of different colors (Cyr et al. 2010) on websites using eye-tracking equipment that measures exactly where and for how long users look at elements of design. Coupled with interviews to determine why users look where they do, these methods offer a systematic analysis of elements of visual design. Most recently, a paper published in the top IS journal *MISQ* by Angelika Dimoka and her colleagues (forthcoming) has charted a research agenda for the use of neurophysiological tools in IS research. The use of methodologies such as eye-tracking and fMRI are part of an evolving research agenda, and are well applicable to visual aesthetics, and the cognitive and affective outcomes for users related to their reactions to visual design principles. In this regard, Soussan Djamasbi (2011) has examined online viewing and aesthetic preferences using an eye-tracking device. These methodologies offer precise insights into why users respond as they do - that serves to develop or elaborate design theory.

### 19.9.3 New Directions for Research

Tractinsky points out important areas for future exploration. I particularly think there is need for additional investigations into how to better serve practitioners through HCI research. Recently a representative of the practitioner community responsible for the production of superior interfaces wrote to me lamenting the need for better design information based on systematic study. Her questions included: why method A is better than method B when running a user evaluation study; how to effectively turn interview data into design criteria for interface development; or how to best determine prototypes that are able to elicit viable evaluation data. This disconnect between designers and users is also outlined by Tractinsky and deserves attention generally, and more specifically in the area of aesthetics and design. Other important areas for investigation outlined in the chapter relate to individual and cultural factors, and as already noted, design differences have been found across cultures for images and color preferences (Cyr et al. 2010; 2011). In a study in which visual design was modeled to trust for Canadians, Germans, and Chinese — only for Chinese users did visual design result in trust (Cyr, 2008). This finding signals diverse reactions to aesthetic elements of websites across different countries. Since work in this area is very sparse, more research is required. The impact of design aesthetics in mobile commerce is also worthy of future investigations. Further, visual design aesthetics significantly impacts perceived usefulness, ease of use, and enjoyment of mobile services (Cyr at al. 2006), representing a novel area for upcoming research in the realm of mobility. Finally, research agendas might explore differences in aesthetics and subsequent reactions to design between men and women. Studies on website design, including visual design, has uncovered significant differences between male and female produced websites (Moss et al. 2006), and related to perceived social presence (i.e. warmth and sociability) of a website as experienced by men versus women (Cyr et al. 2007). While one might expect perceptual differences between men and women to aesthetic stimuli to be well documented, in fact little research resides in this area. In sum, collectively these topics will not only inform and expand theory for aesthetic and visual design, but will provide valuable data for practitioners as well.

### 19.9.4 References


- Cyr, D., Head, M., and Larios, H. (2010). Colour Appeal in Website Design within and across Cultures: A Multi-method Evaluation. *International Journal of Human Computer Studies,* 68(1-2), 1-21.
- Cyr, D., Head, M., Larios, H. and Pan, B. (2009). Exploring Human Images in Website Design: A Multi-Method Approach. *MIS Quarterly*, 33(3), 539-566.
- Cyr, D. (2008). Modeling Website Design across Cultures: Relationships to Trust, Satisfaction and E-loyalty. *Journal of Management Information Systems*, 24(4), 47-72.
- Cyr, D., Hassanein, K., Head, M. and Ivanov, A. (2007). The Role of Social Presence in Establishing Loyalty in e-Service Environments.

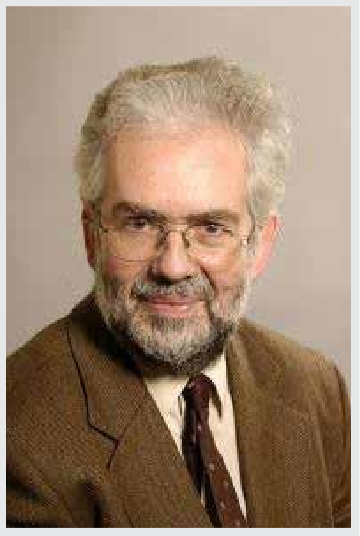
*Interacting with Computers*. Special Issue on “Moving Face-to-Face Communication to Web-based Communication”, 19(1), 43-56.

- Cyr, D., Head, M., and Ivanov, A. (2006). Design Aesthetics Leading to M-loyalty in Mobile Commerce. *Information and Management*, 43(8), 950-963.
- Cyr, D., and Bonanni, C. (2005). Gender and Website Design in E-Business. *International Journal of Electronic Business*, 3(6), 565-582.
- Dimoka, A., Banker, R.D., Benbasat, I., Davis, F.D., Dennis, A.R., Gefen, D., Gupta, A., Ischebeck, A., Kenning, P., Pavlou, P.A., Müller-Putz, G., Riedl, R., vom Brocke, J., and Weber, B. (forthcoming). On the Use of Neurophysiological Tools in IS Research: Developing a Research Agenda for NeuroIS**.** *MIS Quarterly*.
- Djamasbi, S., Siegel, M., Skorinko, J., and Tullis, T. (2011). Online Viewing and Aesthetic Preferences of Generation Y and Baby Boomers: Testing User Website Experience through Eye Tracking, *International Journal of Electronic Commerce*, 15(4), 121-158.
- Gregor, S. and Jones, D. (2007). The Anatomy of a Design Theory. *Journal of the Association for Information Systems*, 8(5), 312-335.
- Moss, G., Gunn, R., and Heller, J. (2006). Some Men Like it Black, Some Women Like it Pink: Consumer Implications of Differences in Male and Female Website Design. *Journal of Consumer Behavior*, 5, 328-341.
- Reimann, M., Zaichkowsky, J., Neuhaus, C., Bender, T. and Weber, B. (2010.) Aesthetic Package Design: A Behavioral, Neural, and Psychological investigation. *Journal of Consumer Psychology*, 20(4), 431-441.

## 19.10 COMMENTARY BY ALISTAIR G. SUTCLIFFE

How to cite this commentary in your report

### Alistair G. Sutcliffe





Alistair Sutcliffe (MA Cantab-Natural Sciences, PhD Wales) is Professor of Systems Engineering, and Director of the Centre for HCI Design, in the School of Informatics, University of Manchester, UK. Originally at ethologist, he has worked in the IT and finance industry, the civil service and City and Manchester Universities. His research spans software engineering, human computer interac...

Alistair G. Sutcliffe
Alistair G. Sutcliffe is a member of The Interaction Design Foundation

Noam Tractinsky has played a key role in placing visual aesthetics on the research agenda of Human Computer Interaction. His 1997 CHI paper launched the phrase 'what is beautiful is usable' to demonstrate that usability was not the only important quality of interactive products, and established aesthetics on the HCI research agenda. As Noam acknowledges, 'what is beautiful is usable' is an adaptation of the beauty in judgement bias well known in psychology, where we attribute more favourable qualities to people we judge to be more handsome or beautiful. However, understanding just how visual aesthetics affects our judgement of products, and how product features influence visual aesthetics, has proven to be a complex and still poorly understood story.

In his chapter Noam examines visual aesthetics from the three perspectives of design, psychology, and pragmatics or practical considerations. Getting designers to define just what constitutes an aesthetic design is a task akin to herding cats; discussion leads to multiple views, perspectives and disagreement. While there have been some attempts to encapsulate good principles of aesthetic design (e.g. Kristof & Satran 1995, Lidwell et al., 2002), design is a highly creative activity which can never be formally analysed, so aesthetic design continues to expand into new frontiers. In his chapter Noam tries to restrict himself to visual aesthetics rather than reviewing the wider area of user experience (UX) which involves other product qualities such as interactive features, customisation and adaptation, as well as content and services. Limiting discussion to visual aesthetics may keep a chapter within page limits, but it is difficult to draw the line between visual aesthetics and user experience; for example, is our reaction to an interactive animated character determined by its appearance (visual aesthetics), how it interacts or a combination of both? In my own work, with Antonella de Angeli, we have been trying to tease apart a multi-faceted view of product quality judgement of which visual aesthetic is but one component (De Angeli et al., 2006; Hartman et al., 2007, 2008).

From the psychology perspective, Noam laid the foundations for quantitative measurement of visual aesthetics with his classic and expressive aesthetic scales (Lavie & Tractinsky, 2004), later expanded with symbolic and pleasure scales. I have used these questionnaires many times to explore the beauty and usability debate, showing that the initial 'what is beautiful is usable' was a bit of a simplification. In fact the 'halo' effect, where favourable judgement on one quality (aesthetics) will spill over into another (usability), is highly context-dependent and users' judgement on the same product will change dramatically according to the task and between users (Sutcliffe, 2009; Hartmann et al., 2007, 2008). Overall judgement about product quality appears to be a complex interaction between several qualities: content/services, visual aesthetics, interaction, customisation, and product identity/brand. Furthermore, as Noam notes, judgement of aesthetics changes over time, from initial almost subliminal impressions, first demonstrated by Gitte Lindgaard, to more reflective and cognitive assessment of quality. This may explain why getting a consensus about a design from designers is a lost cause: not only is beauty in the eye of the beholder, it also changes over time.

I use the term 'user engagement' to cover not only visual aesthetics but also interactive qualities of products which can range from simple menu-link navigation on websites to 3D graphical worlds with interactive avatars (virtual people) and flying through navigation as found in SecondLife, World of Warcraft and a host of games applications. While visual aesthetics is important at first sight, interactivity and functionality soon become much more important, as users' judgement changes within a session and over successive encounters (Hartman et al., 2008; Sutcliffe, 2009). The essence of user engagement is illustrated in Figure 19.1.

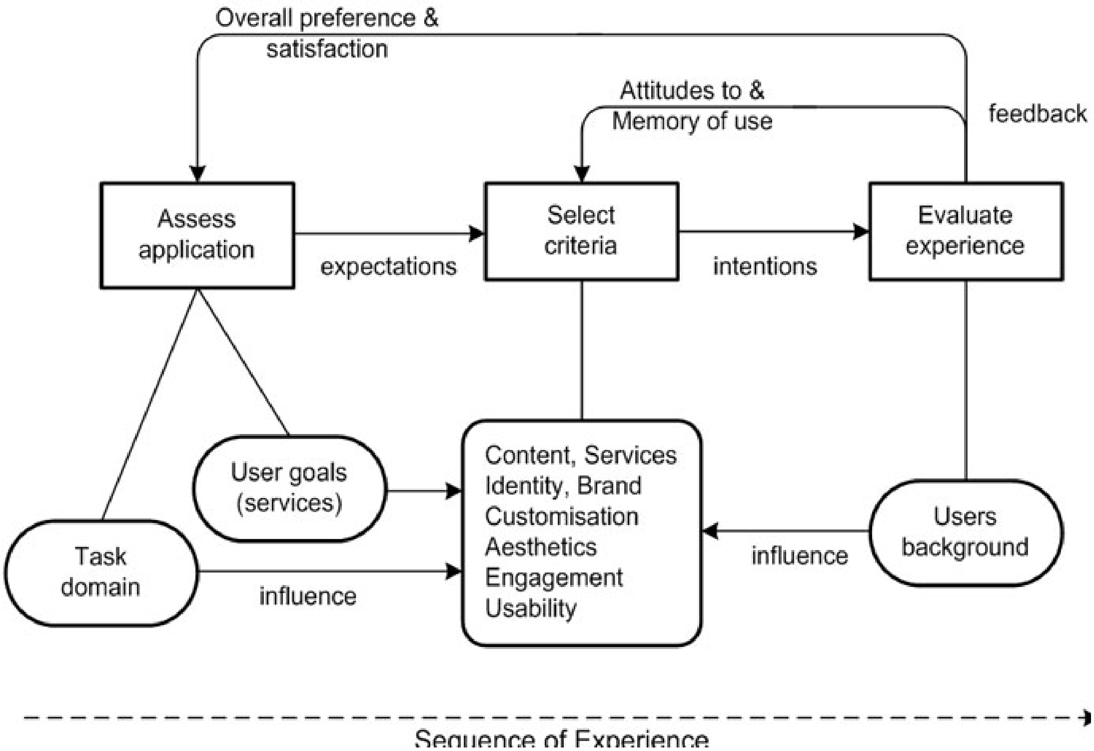



Figure 19.1: Model of user engagement, showing the interplay between judgement criteria and the user-domain context.



Not only do the criteria influencing user judgement change over time, but they also depend on the product domain. Our preliminary theory predicts that as users experience progresses with more interactive sessions the criteria which are important for judgement change. On first sight aesthetics is important but then interaction and engagement takes over, however in the longer run utility (content and services) become dominant. Usability must be good enough so it doesn't annoy the user but not perfect- people will forgive small problems. The application domain also biases the criteria. For games, interactivity and flow are paramount, but would you want to do your online banking in SecondLife? Well, maybe some of you would. Discovering where the general laws of user quality judgement and preferences lie will keep myself, Noam and many others occupied for many years to come.

The psychology and pragmatic perspectives may have a closer relationship than is immediately apparent. Noam reviews Marc Hassenzahl's work in his chapter, describing the hedonic (pleasure/aesthetic) and pragmatic (usability/utility) constructs which Marc has shown to be remarkably consistent over a range of products and users (Hassenzahl, 2004, 2010). Since hedonic and pragmatic constructs are related to simpler concepts of goodness and beauty, maybe we judge products by these two simple constructs; or, as Noam and I believe, the picture is more complex with components such as classic and expressive aesthetics, user engagement, service quality, etc., competing to compose the final impression of satisfaction, emotional reaction or preference. All of us believe judgement is context-dependent, but there is disagreement about where the boundary of general psychology and the influence of context lies. Noam failed to mention in his chapter the contextual school of user experience, headed by John McCarthy and Peter Wright (McCarthy & Wright, 2004, 2010), who hold the view that all user experience can only be understood by investigating the 'dialogue' or co-experience between the user, product and context, aided by interpretivist theory and a qualitative, epistemological approach. The contextualists, and indeed many designers, would argue that visual aesthetics is a pragmatic endeavour which can only be analysed in context and created through experience. I think Noam disagrees with this view and will continue, as will I, to unpack the psychology of visual aesthetics and user experience, and may one day even relate our measures to design principles and features. However, I suspect we will never catch up with the designers; the motor of creative aesthetics runs faster than the process of scientific research.

### 19.10.1 References


- De Angeli, A., Sutcliffe, A.G. & Hartmann, J. (2006). Interaction, usability and aesthetics: What influences users' preferences? *Proceedings: Conference on Designing Interactive Systems, DIS-06*, pp. 271-280. New York: ACM Press.

- Hartmann, J., Sutcliffe, A.G., & De Angeli, A. (2007). Investigating attractiveness in web user interfaces. *Proceedings: CHI-07, San Jose*. New York: ACM Press.
- Hartmann, J., Sutcliffe, A.G., and De Angeli, A. (2008). Towards a theory of user judgement of aesthetics and user interface quality. *ACM Transactions on Computer-Human Interaction*, 15(4).
- Hassenzahl, M. (2004). The interplay of beauty, goodness and usability in interactive products. *Human-Computer Interaction*, 19, 319-349.
- Hassenzahl, M. (2010). Experience design: Technology for all the right reasons. In Carroll, J.M. (Ed.) *Synthesis lectures on human centered informatics*. San Rafael CA: Morgan Claypool.
- Kristof, R. & Satran, A. (1995). *Interactivity by design: Creating and communicating with new media*. Mountain View CA: Adobe Press.
- Lavie, T. & Tractinsky, N. (2004). Assessing dimensions of perceived visual aesthetics of web sites. *International Journal of Human-Computer Studies*, 60, 269-298.
- Lidwell, W., Holden K. & Butler, J. (2003). *Universal principles of design*. Gloucester MA: Rockport.
- McCarthy, J. & Wright, P. (2005). *Technology as experience*. Cambridge MA: MIT Press.
- McCarthy, J. & Wright, P. (2010). Experience centred design: Designers, users and communities in dialogue. In Carroll, J.M. (Ed.) *Synthesis lectures on human centered informatics*. San Rafael CA: Morgan Claypool.
- Sutcliffe, A.G. (2009). Designing for user engagement: aesthetic and attractive user interfaces. In Carroll, J.M. (Ed.) *Synthesis lectures on human centered informatics*. San Rafael CA: Morgan Claypool.

## 19.11 COMMENTARY BY JINWOO KIM

How to cite this commentary in your report

**Jinwoo Kim**

*© Jinwoo Kim*

Professor at Yonsei University, Korea, and and director of the HCI Lab at Yonsei. PhD from Carnegie Mellon University....

Jinwoo Kim
Jinwoo Kim is a member of The Interaction Design Foundation

“Do we have to study this stuff? This is so ... obvious...”

This is a comment that I got from my graduate student about ten years ago. At that time, a few groundbreaking research outputs were released such as (Kurosu and Kashimura, 1995) and (Tractinsky et al., 2000). As a half-cooked cognitive engineer with a strong inclination to empirical validation, I was fascinated with these papers, and I decided to use them as discussion material for my graduate-level HCI class in which about one-third of the students were from the Korean design industry. However, to them, the fact that visual aesthetics were closely related to the overall quality of system use was so obvious that they did not feel any need for serious research. However, it is equally surprising to me that not much had been known about the importance, the antecedents and outcomes, and the moderating conditions of 'visual aesthetics' until these few groundbreaking studies. And I believe this is the core message of the article 'visual aesthetics' written by Noam Tractinsky.

The importance of visual aesthetics has also been found in one of our studies that applied the three Viturvian design principles (mentioned in Noam's article) on four different kinds of Internet businesses (Kim et al., 2002). We found that venustas (or visual aesthetics) had strong influences on user satisfaction and customer loyalty. An even more interesting finding was that visual aesthetics influenced user satisfaction even in intrinsically utilitarian domains such as online stock brokerages and search portals.

I would like to add two more reasons for why 'visual aesthetics' will become even more important in the future.

*First*, user experience (UX) will get more attention in the future. UX is defined as "a person's perceptions and responses that result from the use or anticipated use of a product, system or service (ISO 9241-210)." One of the main characteristics of UX, in comparison to usability, is that it is more subjective and holistic. In relation, visual aesthetics has been found to strongly affect subjective and holistic experience (Park et al., 2005). Corroborating evidence for this prediction comes from the broader participation of design communities in the HCI field as UX becomes more important. For example, a recent annual meeting of the Korean HCI society attracted more than 1,500 participants where more than one-third of papers at the conference came from designers who were interested in UX.

*Second*, as information technologies (such as smart phones and tablet computers) become more ubiquitous, the quality of life, not just use quality, has been greatly influenced. One of our studies (Choi et al. 2007) investigated the impact of use experience upon the overall quality of life. The results indicate that several life domains such as cultural and leisure as well as financial and educational domains are greatly affected by the visual characteristics of mobile technologies. Visual aesthetics will hence have greater impact on our quality of life as information technologies are utilized more pervasively.

In order to meet the growing importance of visual aesthetics, I would like to add three more future research directions.

*First*, most prior studies in visual aesthetics focused on individual user experience such as a single person using a web page or mp3 players to evaluate visual aesthetics felt only by him/her. However, as social computing is used more pervasively, more people are using products and services together. For example, people use YouTube and leave their comments on the visual aesthetics of video content. As a matter of fact, comments and opinions on visual aesthetics are more frequently observed than those on usefulness or usability. However, not much has been known about how people express their visual aesthetics to the public and how other people are affected by comments and opinions on visual aesthetics. Future studies should investigate the social formation process of visual aesthetics.

*Second*, most prior studies in visual aesthetics have investigated the consumption process of visual aesthetics, but not much research has been conducted on the creation process of visual aesthetics by ordinary users. However, a recent phenomenon reveals average users who create artifacts that focus on visual aesthetic properties. For example, visual aesthetics are main focus of T-shirts at Threadless.com and appliances at Quirky.com. However, most research in user creation focused on the utilitarian and economic perspectives by expert users, and not many studies have been conducted on the creation process of visual aesthetics by ordinary users. Future studies on the creation process of lay users will contribute not only to visual aesthetics research but to user innovation research as well. These studies also provide greater practical implications for facilitating the creation process of visually appealing outputs by lay users.

*Third*, current interests in service design may prompt an interesting question on visual aesthetics for non-visual artifacts. Most prior studies on visual aesthetics have focused on some tangible and visible artifacts such mobile phone skins or ATM machines. However, we observe that people frequently mention the visual aesthetic aspects of services such as shopping or searching. As was mentioned in Noam's article, not much attention was paid to dynamic aspects of visual aesthetics. How do people experience visual aesthetics for intangible services and dynamic contents? Do they feel visual aesthetics from the visual components of the services or are they affected by some neurological stimuli that mediate between non-visible aspects of services and visual aesthetics? These are the questions that might provide additional explanations on visual aesthetic processes.

In summary, visual aesthetics will be more important as IT products and services become ubiquitous and holistic. Noam's framework on the antecedents, evaluation process, moderators and outcomes of visual aesthetics in HCI will provide valuable starting points for us to understand more deeply the intuitively obvious but poorly investigated visual aesthetics.

### 19.11.1 References


- Choi, H., Lee, M., Im, K., and Kim, J. (2007). Contribution to Quality of Life: A New Outcome Variable for Mobile Data Service, Journal of the Association for Information Systems, 8 (12), 598-618.
- Kim, Jinwoo, Lee, Jungwon, Han, Kwanghee and Lee, Moonkyu (2002): Businesses as Buildings: Metrics for the Architectural Quality of Internet Businesses. In Information Systems Research, 13 (3) pp. 239-254
- Kurosu, Masaaki and Kashimura, Kaori (1995): Apparent usability vs. inherent usability: experimental analysis on the determinants of the apparent usability. In: CHI 95 Conference Companion 1995 1995. pp. 292-293.
- Park, Su-e, Choi, Dongsung and Kim, Jinwoo (2005): Visualizing E-Brand Personality: Exploratory Studies on Visual Attributes and E-Brand Personalities in Korea. In International Journal of Human-Computer Interaction, 19 (1) pp. 7-34.

- Tractinsky, Noam, Katz, A. S. and Ikar, D. (2000): What is Beautiful is Usable. In Interacting with Computers, 13 (2) pp. 127-145.

## 19.12 COMMENTARY BY MASAAKI KUROSU

How to cite this commentary in your report

### Masaaki Kurosu

© *Masaaki Kurosu*

Masaaki Kurosu is a professor of the R&D Department of National Institute of Multimedia Education (NIME) of MEXT, the ministry of education. He is also a head of Department of Cyber Culture and Society of the Graduate University for Advanced Studies (SOKENDAI). Before coming to NIME and SOKENDAI on September of 2001, he was a professor at the Faculty of Information of Shizuoka University...

Masaaki Kurosu
Masaaki Kurosu is a member of The Interaction Design Foundation

### 19.12.1 Aesthetics and Beauty

First, I would like to use the word "beauty" instead of "aesthetics" unlike Tractinsky for the reason that the former is a quality characteristics of the object while the latter is a philosophical consideration on the beauty. And the latter is rooted in Western culture since Greek era (more specifically since Bavmgarten in 1750) and is thus specific to Western culture, while the concept of beauty is universal although its connotation and denotation varies so much depending on time and culture.

In Japan, for example, craftsman who made unglazed earthenware and drew patterns on its surface in ancient times must have some intension for the beauty and, possibly, for some religious significance.

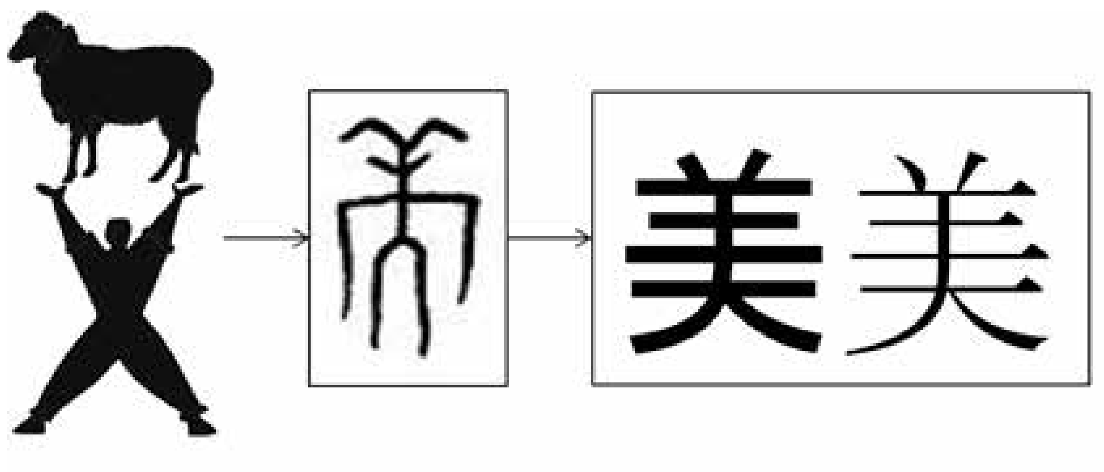



Figure 19.1: The generation and development of Chinese character meaning the beauty.



The conceptualization of beauty in Japan was influenced by China at the time of import of Chinese characters at 5-6th century. As is shown in Figure 19.1, the Chinese character meaning the beauty consists of the image of a sheep and a man, thus is describing some religious and ritual meaning. After obtaining the Chinese character, the concept of beauty could have been externalized in Japan and since then many beautiful Japanese arts and crafts have been made though not having the philosophical consideration on the nature of beautifulness but with the intension for the beauty.

In 1875, early in Meiji era when Japan stopped her national isolation and started to mass-import Western culture, Nishi translated the concept of "Aesthetica" of Bavmgarten and used the term "美學" that is still used now for translating the word "aesthetics". It was the starting point of aesthetics in Japan. In other words, Japanese had a concept of beauty for quite a long time but the concept of aesthetics for only 150 years.

## 19.12.2 Beauty and Art

In 1881, Fiedler, the originator of "Kunstwissenschaft", made the science of art to be separated from the aesthetics and regarded the substance of art independent from the beauty. There are so many examples of work of art that are not "beautiful" including "Les Masques et La Mort" by Ensor (1897), "Fountain" by Duchamp (1917), "Die Frauen der Revolution" by Kiefer (1987), etc.

FIGURE 19.2: Artworks that are not beautiful: "Les Masques et La Mort" by Ensor, J. (1897), "Fountain" by Duchamp, M. (1917) and "Die Frauen der Revolution" by Kiefer, A. (1987).

In between the art and design, there is another example of advertisement by Benetton as shown in Figure 19.3.

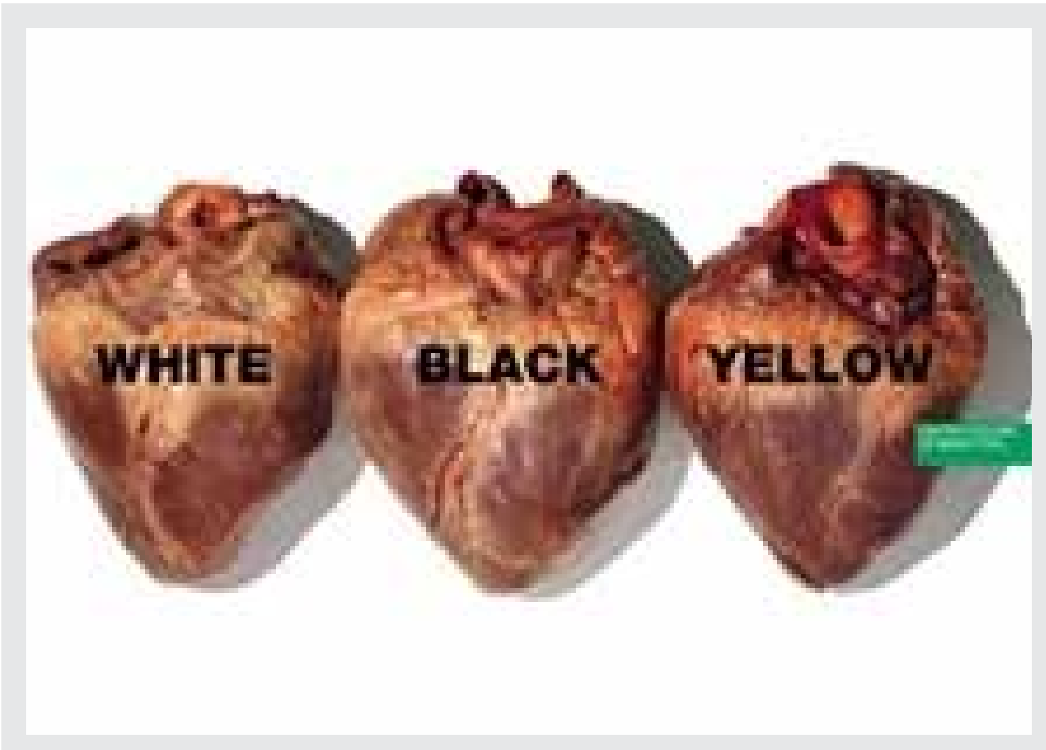



Figure 19.3: Advertisement by Benetton (Photo: Oliviero Toscani).



Of course, there are many art works today that are beautiful, but it is also true that there exist other artists who want to present their work so that viewers will consider such serious themes as the meaning of life, the relationship between people and object, the peace and war, the human rights, etc.

### 19.12.3 Beauty and Design

Designing is a universal human activity that can be found anytime and anywhere. The key point that the design is different from the art is that there is a user for the design. In most cases, users are other people than designers and designed products will be merchandised. As a result, designers make efforts to let their output be attractive to users. And one of the key elements of this attractiveness is the beauty.

Judgment on the designed product is not much complex compared to the judgment on the work of art maybe because it is related more to the perceptual process than to the conceptual process, especially in terms of the beauty. Hence, as Tractinsky pointed out, the law of symmetry, the law of simplicity, the law of grid design, etc. can be applied and be perceived to increase the degree of beauty of designed product.

But a simple application of such laws of designing can generate difficult-to-use products as shown in Figure 19.4 and Figure 19.5.

FIGURE 19.4: Beautiful but difficult-to-use design (example 1) - A misuse of the law of symmetry.

In Figure 19.4, the UI layout of the laptop is shown where the touch pad is placed at the center of the body by applying the law of symmetry. The designer, thus, neglected the touch typing usability. As is well known, the touch typing for the fast text input requires four fingers of each hand to be placed on the home positions; "asdf" for the left hand and "jkl;" for the right hand. But if you try to place your hand on this keyboard, you will find that the palm of the right hand will be placed on the touch pad and unintended cursor movement will occur (lower left picture). And if you try to avoid unexpected touching to the pad, you will have to put your hands in an awkward manner (lower right picture). If the designer follows the law of usability (in general), the location of the touch pad should be displaced a bit to the left.

Figure 19.5: Beautiful but difficult-to-use design (example 2) - Overemphasizing the color design.

In Figure 19.5, a calculator is shown that looks beautiful regarding the color design. But as you can see in the lower right picture, the assignment of numbers and symbols are quite difficult to see because of the low contrast between the figure and the ground. For the designer, I guess, numbers and symbols with high contrast to the background were just the visual noise. Thus s/he might have violated the law of usability (in general).

### 19.12.4 Beauty, Quality Characteristics and Meaning in Design

As was pointed out by Tractinsky, the visual judgment on beauty is very fast, thus plays an important role in drawing the attention of customers. And the visual beauty is dominated by rather simple and traditional rules. But too much emphasis on the beauty will lead to a difficult-to-use designs as was discussed in the previous section. Hence the usability or the pragmatic aspects is important at the same level as the pleasure or the hedonic attributes as Jordan (1999) and Hassenzahl (2003) pointed out.

Designers and marketing people have a tendency to put more emphasis on the attractiveness of the product, thus tend to focus on the beauty, pleasure and hedonic aspects. But it is only a one-sided approach. We should remember that the consumer will become the user after the purchase of the product and will start using it. Unlike designers and marketing people, usability professionals, ergonomics specialists and engineers tend to put a bit too much emphasis on the phase of the user and focus on the usability and functionality. Although this is another type of approach, two types of stakeholders will have to cooperate in a well-balanced manner based on the understanding of the result of Kurosu and Kashimura (1995) that the apparent usability will not cover the inherent usability.

| Manufacturer | Market | Kurosu, M. and Kashimura, K. (1995) | Jordan, P. (1999) | Hassenzahl, M. (2003) | Kurosu, M. (2012) |
|---|---|---|---|---|---|
| Designers and Marketing People | Consumer | Beauty (Apparent Usability) | Pleasure | Hedonic Attributes | Subjective Quality (To be felt and to be attracted) |
| Usability Professionals and Engineers | User | Inherent Usability | Usability Functionality | Pragmatic Attributes | Objective Quality (To be judged) |
| | | | | | Meaningfulness (To be accepted) |

Figure 19.6: Three Dimensions of Design.



Figure 19.6 summarizes this discussion in terms of the cooperative design between two disciplined stakeholders. Furthermore, this table shows the importance of the meaningfulness as was proposed by Kurosu (2012) in addition to the subjective quality characteristics (beauty, pleasure or hedonic attributes) and the objective quality characteristics (usability, functionality, performance, reliability, safety, maintenability, etc.)

The meaningfulness in the third row of this table means to design what people really needs. A typical example appeared in the Japanese market recently: a television set equipped with the ionized air emitting function that will allow users to watch the program in a good physical environment was released. Is this what people needed? Should these functions be united together?

Even if a product is attractively designed and have an acceptable level of objective quality, that product will be useless if it doesn't have a meaning. This is the reason why Kurosu added the meaningfulness to the subjective quality and the objective quality.

It should be admitted that the beauty as one of the key subjective quality characteristics is quite important. But taking a good balance among these three dimensions should not be forgotten.

## 19.12.5 References


- Bavmgarten, A.G. (1750) “Aesthetica”
- http://books.google.co.jp/books?id=iPVNLFAGznQC&printsec=frontcover&hl=ja&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false
- Fiedler, K.A. (1881) “Moderner Naturalismus und künstlerische Wahrheit”
- Hassenzahl, M. (2003) “The Thing and I: Understanding the Relationship Between User and Product” in Blythe, M., Overbeeke, C., Monk, A.F., and Wright, P.C. (eds.) “Funology: From Usability to Enjoyment” Kluwer, pp.31-42
- Jordan, P.W. (1999) “Pleasure with Products: Human Factors for Body, Mind and Soul” in Green, W.S. and Jordan, P.W. (eds.) “Human Factors in Product Design” Taylor & Francis
- Kurosu, M. (2012). Three Dimensions of Artifact Design — Meaningfulness, QualityTraits and Kansei (in Japanese). Human Interface Symposium 2012
- Kurosu, M. and Kashimura, K. (1995): Apparent usability vs. inherent usability: experimental analysis on the determinants of the apparent usability. In: CHI 95 Conference Companion 1995 1995. pp. 292-293
- Nishi, A. (1875) “Â Â” (in Japanese)
- Tractinsky, N. (1997): Aesthetics and apparent usability: empirically assessing cultural and methodological issues. In: Proceedings of the SIGCHI conference on Human factors in computing systems 1997. pp. 115-122

## 19.13 REFERENCES


*Angeli, Antonella De, Sutcliffe, Alistair G. and Hartmann, Jan (2006): Interaction, usability and aesthetics: what influences users' preferences?. In: Proceedings of DIS06: Designing Interactive Systems: Processes, Practices, Methods, & Techniques 2006. pp. 271-280*

*Arnheim, Rudolf (1966): Order and complexity in landscape design. In: Arnheim, Rudolf (ed.). "Toward a Psychology of Art". University of California Press*

*Bargas-Avila, Javier A. and Hornbæk, Kasper (2011): Old wine in new bottles or novel challenges: a critical analysis of empirical studies of user experience. In: Proceedings of ACM CHI 2011 Conference on Human Factors in Computing Systems 2011. pp. 2689-2698*

*Bateman, Scott, Mandryk, Regan L., Gutwin, Carl, Genest, Aaron, McDine, David and Brooks, Christopher (2010): Useful junk?: the effects of visual embellishment on comprehension and memorability of charts. In: Proceedings of ACM CHI 2010 Conference on Human Factors in Computing Systems 2010. pp. 2573-2582*

*Bauerly, Michael and Liu, Yili (2006): Computational modeling and experimental investigation of effects of compositional elements on interface and design aesthetics. In International Journal of Human-Computer Studies, 64 (8) pp. 670-682*

*Belk, Russell W. (1988): Possessions and the Extended Self. In Journal of Consumer Research, 15 (2) pp. 139-168*

*Ben-Bassat, Tamar, Meyer, Joachim and Tractinsky, Noam (2006): Economic and subjective measures of the perceived value of aesthetics and usability. In ACM Transactions on Computer-Human Interaction, 13 (2) pp. 210-234*

*Bergkvist, Lars and Rossiter, John R (2007): The Predictive Validity of Multiple-Item Versus Single-Item Measures of the Same Constructs. In Journal of Marketing Research, 44 (2) pp. 175-184*

*Berlyne, D. E. (1971): Aesthetics and psychobiology. Appleton-Century-Crofts*

*Bloch, Peter H. (2011): Product Design and Marketing: Reflections After Fifteen Years. In Design, 28 (3) pp. 378-380*

*Bloch, Peter H., Brunel, F. F. and Arnold, T. J. (2003): Individual Differences in the Centrality of Visual Product Aesthetics: Concept and Measurement. In Journal of Consumer Research, 29 (4) pp. 551-565*

*Brooks, Fred (1975): The Mythical Man-Month: Essays on Software Engineering. Addison-Wesley Publishing*

Cai, Shun and Xu, Yunjie (Calvin) (2011): Designing Not Just for Pleasure: Effects of Web Site Aesthetics on Consumer Shopping Value. In International Journal of Electronic Commerce, 15 (4) p. 159–187

Card, Stuart K., Moran, Thomas P. and Newell, Allen (1983): The Psychology of Human-Computer Interaction. Hillsdale, NJ, Lawrence Erlbaum Associates

Carlson, Allen (2000): Aesthetics and the Environment: The Appreciation of Nature, Art and Architecture. Routledge

Cawthon, Nick and Moere, Andrew Vande (2007): The Effect of Aesthetic on the Usability of Data Visualization. In: Proceedings of the 11th International Conference Information Visualization 2007. pp. 637-648

Chen, Chaomei (2005): Top 10 unsolved information visualization problems. In IEEE Computer Graphics and Applications, 25 (4) pp. 12-16

Creusen, Marielle E. H. and Schoormans, Jan P. L. (2005): The Different Roles of Product Appearance in Consumer Choice. In Journal of Product Innovation Management, 22 (1) pp. 63-81

Csikszentmihalyi, Mihaly (1991): Design and Order in Everyday Life. In Design Issues, 8 (1) pp. 26-34

Cyr, Dianne (2008): Modeling Web Site Design Across Cultures: Relationships to Trust, Satisfaction, and E-Loyalty. In Journal of Management Information Systems, 24 (4) pp. 47-72

Cyr, Dianne, Head, Milena M. and Ivanov, Alex (2006): Design aesthetics leading to m-loyalty in mobile commerce. In Information & Management, 43 (8) pp. 950-963

Cyr, Dianne, Kindra, Gurprit S. and Dash, Satyabhusan (2008): Web site design, trust, satisfaction and e-loyalty: the Indian experience. In Online Information Review, 32 (6) pp. 773-790

Cyr, Dianne, Head, Milena and Larios, Hector (2010): Colour appeal in website design within and across cultures: A multi-method evaluation. In International Journal of Human-Computer Studies, 68 (1) pp. 1-21

Datta, Ritendra, Li, Jia and Wang, J. Z. (2008): Algorithmic inferencing of aesthetics and emotion in natural images: An exposition. In: Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on 2008. pp. 105-108

Datta, Ritendra, Joshi, D., Li, J. and Wang, James Z. (2006): Studying aesthetics in photographic images using a computational approach. In Distribution, 3953 pp. 288-301

*Davis, Fred D. (1989): Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. In MIS Quarterly, 13 (3) pp. 319-340*

*Dion, Karen, Berscheid, Ellen and Walster, Elaine (1972): What is beautiful is good. In Journal of Personality and Social Psychology, 24 (3) pp. 285-290*

*Dutton, Denis (2008): The Art Instinct: Beauty, Pleasure, and Human Evolution. Bloomsbury Press*

*Eagly, Alice H., Ashmore, Richard D., Makhijani, Mona G. and Longo, Laura C. (1991): What is beautiful is good, but..: A meta-analytic review of research on the physical attractiveness stereotype. In Psychological Bulletin, 110 (1) pp. 109-128*

*Fishwick, Paul A. (ed.) (2006): Aesthetic Computing. The MIT Press*

*Forlizzi, Jodi (2008): The product ecology: Understanding social product use and supporting design culture. In International Journal of Design, 2 (1) pp. 11-20*

*Gardner, D. G., Cummings, L. L., Dunham, R. B. and Pierce, J. L. (1998): Single-Item Versus Multiple-Item Measurement Scales: An Empirical Comparison. In Educational and Psychological Measurement, 58 (6) pp. 898-915*

*Gifford, Robert, Hine, Donald W, Muller-Clemm, Werner, Reynolds, D'arcy J. and Shaw, Kelly T. (2000): Decoding Modern Architecture: A Lens Model Approach for Understanding the Aesthetic Differences of Architects and Laypersons. In Environment And Behavior, 32 (2) pp. 163-187*

*Gladwell, Malcolm (2000): The Tipping Point: How Little Things Can Make a Big Difference. Back Bay Books*

*Hamermesh, Daniel S. and Biddle, Jeff E. (1994): Beauty and the Labor Market. In American Economic Review, 84 (5) pp. 1174-1194*

*Hamermesh, Daniel S. and Parker, Amy M. (2003): Beauty in the Classroom: Professors' Pulchritude and Putative Pedagogical Productivity. In National Bureau of Economic Research Working Paper Series, 9853 (4) pp. 369-376*

*Hartmann, Jan, Sutcliffe, Alistair G. and Angeli, Antonella De (2008): Towards a theory of user judgment of aesthetics and user interface quality. In ACM Transactions on Computer-Human Interaction, 15 (4) p. 15*

*Hartmann, Jan, Sutcliffe, Alistair G. and Angeli, Antonella De (2007): Investigating attractiveness in web user interfaces. In: Proceedings of ACM CHI 2007 Conference on Human Factors in Computing Systems 2007. pp. 387-396*

*Hassenzahl, Marc (2007): Aesthetics in interactive products: Correlates and consequences of beauty. In: Schifferstein, Hendrik N. J. and Hekkert, Paul (eds.). "Product Experience". Elsevier Science*

*Hassenzahl, Marc (2003): The Thing and I: Understanding the Relationship Between User and Product. In: Blythe, Mark.A., Overbeeke, Kees, Monk, Andrew F. and Wright, Peter C. (eds.). "Funology: From Usability to Enjoyment". Springerpp. 31-42*

*Hassenzahl, Marc (2004b): Beautiful Objects as an Extension of the Self: A Reply. In Human-Computer Interaction, 19 (4) pp. 377-386*

*Hassenzahl, Marc (2004a): The Interplay of Beauty, Goodness, and Usability in Interactive Products. In Human-Computer Interaction, 19 (4) pp. 319-349*

*Hassenzahl, Marc and Monk, Andrew (2010): The Inference of Perceived Usability From Beauty. In Human Computer Interaction, 25 (3) pp. 235-260*

*Hassenzahl, Marc and Tractinsky, Noam (2006): User experience - a research agenda. In Behaviour and Information Technology, 25 (2) pp. 91-97*

*Heijden, Hans van der (2003): Factors influencing the usage of websites: the case of a generic portal in The Netherlands. In Information and Management, 40 (6) pp. 541-549*

*Hekkert, Paul, Snelders, Dirk and Wieringen, Piet C. W. (2010): "Most advanced, yet acceptable": Typicality and novelty as joint predictors of aesthetic preference in industrial design. In British Journal of Psychology, 94 (1) pp. 111-124*

*Hooper, Kristina (1986): Architectural design: an analogy. In: Norman, Donald A. and Draper, Stephen W. (eds.). "User Centered System Design: New Perspectives on Human-Computer Interaction". Hillsdale, NJ: Lawrence Erlbaum Associates*

*Hoyer, Wayne and Stokburger-Sauer, Nicola (2011): The role of aesthetic taste in consumer behavior. In Journal of the Academy of Marketing Science,*

*Höfela, Lea and Jacobsena, Thomas (2007): Electrophysiological Indices of Processing Symmetry and Aesthetics: A Result of Judgment Categorization or Judgment Report?. In Journal of Psychophysiology, 21 (1) pp. 9-21*

*Inbar, Ohad, Tractinsky, Noam and Meyer, Joachim (2007): Minimalism in information visualization: attitudes towards maximizing the data-ink ratio. In: Brinkman, Willem-Paul, Ham, Dong-Han and Wong, B. L. William (eds.) ECCE 2007 - Proceedings of the 14th European Conference on Cognitive Ergonomics August 28-31, 2007, London, UK. pp. 185-188*

*Jacobsen, Thomas (2004): Individual and group modelling of aesthetic judgment strategies. In British journal of psychology London England 1953, 95 (0) pp. 41-56*

*Jacobsen, Thomas (2010): Beauty and the brain: culture, history and individual differences in aesthetic appreciation. In Journal of Anatomy, 216 (2) pp. 184-191*

*Johnson, Paul-Alan (1994): The Theory of Architecture: Concepts, Themes & Practices. Van Nostrand Reinhold*

*Jones, Edward Ellsworth (1990): Interpersonal Perception. W H Freeman and Co (Sd)*

*Jr., Frank Gibney and Luscombe, Belinda (2008). The Redesigning Of America. Retrieved 1 December 2011 from Times.com: http://www.time.com/time/magazine/article/0,9171,996372,00.html*

*Karapanos, Evangelos, Zimmerman, John, Forlizzi, Jodi and Martens, Jean-Bernard (2010): Measuring the dynamics of remembered experience over time. In Interacting with Computers, 22 (5) pp. 328-335*

*Kim, Jinwoo and Moon, Jae Yun (1998): Designing Towards Emotional Usability in Customer Interfaces -- Trustworthiness of Cyber-Banking System Interfaces. In Interacting with Computers, 10 (1) pp. 1-29*

*Kim, Jinwoo, Lee, Jungwon, Han, Kwanghee and Lee, Moonkyu (2002): Businesses as Buildings: Metrics for the Architectural Quality of Internet Businesses. In Information Systems Research, 13 (3) pp. 239-254*

*Kim, Jinwoo, Lee, Jooeun and Choi, Dongseong (2003): Designing emotionally evocative homepages: an empirical study of the quantitative relations between design factors and emotional dimensions. In International Journal of Human-Computer Studies, 59 (6) pp. 899-940*

*Kleine, Susan Schultz, Kleine, Robert E. and Allen, Chris T. (1995): How Is a Possession "Me" or "Not Me"? Characterizing Types and an Antecedent of Material Possession Attachment. In Journal of Consumer Research, 22 (3) pp. 327-43*

*Korman-Golander, Gili (2011). Aspects of Fashion/Trends in Web Design, M.Sc. Thesis. Ben-Gurion Univerity of the Negev http://aranne5.lib.ad.bgu.ac.il/others/KormanGolander-Gili.pdf*

*Krippendorff, Klaus (2005): The Semantic Turn: A New Foundation for Design. CRC Press*

*Kruft, Hanno-Walter (1994): A History of Architectural Theory: From Vitruvius to the Present. Zwemmer*

*Kumara, Minu and Gargb, Nitika (2010): Aesthetic principles and cognitive emotion appraisals: How much of the beauty lies in the eye of the beholder?. In Journal of Consumer*

*Psychology, 20 (4) pp. 485-494*

Kurosu, *Masaaki and* Kashimura, *Kaori (1995): Apparent usability vs. inherent usability: experimental analysis on the determinants of the apparent usability. In: CHI 95 Conference Companion 1995 1995. pp. 292-293*

Lavie, Talia *and* Tractinsky, *Noam (2004): Assessing dimensions of perceived visual aesthetics of web sites. In International Journal of Human-Computer Studies, 60 (3) pp. 269-298*

Law, Effie L.-C. *and* Schaik, *Paul Van (2010): Modelling user experience -- An agenda for research and practice. In Interacting with Computers, 22 (5) pp. 313-322*

Leder, Helmut, Belke, Benno, Oeberst, *Andries and* Augustin, *Dorothee (2004): A model of aesthetic appreciation and aesthetic judgments. In British Journal of Psychology, 95 (0) pp. 489-508*

Lee, *Allen S. (1991): Architecture as a reference discipline for MIS. In:* Nissen, *H. E.,* Klein, *H. K. and* Hirschheim, *R. (eds.). "Information Systems Research: Contemporary Approaches & Emergent Traditions". North Hollandp. 573–592.*

Lee, *Sangwon and* Koubek, *Richard J. (2010): Understanding user preferences based on usability and aesthetics before and after actual use. In Interacting with Computers, 22 (6) pp. 530-543*

Lindgaard, Gitte *and* Dudek, *Cathy (2003): What is this evasive beast we call user satisfaction?. In Interacting with Computers, 15 (3) pp. 429-452*

Lindgaard, Gitte, Dudek, *Cathy,* Sen, *Devjani,* Sumegi, *Livia and* Noonan, *Patrick (2011): An exploration of relations between visual appeal, trustworthiness and perceived usability of homepages. In ACM Transactions on Computer-Human Interaction (TOCHI), 18 (1)*

Lindgaard, Gitte, Fernandes, *Gary,* Dudek, *Cathy and* Brown, *J. (2006): Attention web designers: You have 50 milliseconds to make a good first impression. In Behaviour and Information Technology, 25 (2) pp. 115-126*

Liu, *Yili (2003): Engineering aesthetics and aesthetic ergonomics: theoretical foundations and a dual-process research methodology. In Ergonomics, 46 (13) pp. 1273-1292*

Luchs, *Michael and* Swan, *K. Scott (2011): Perspective: The Emergence of Product Design as a Field of Marketing Inquiry. In Journal of Product Innovation Management, 28 (3) pp. 327-345*

Lyubomirsky, *Sonja,* King, *Laura and* Diener, *Ed (2005): The benefits of frequent positive affect: does happiness lead to success?. In Psychological Bulletin, 131 (6)*

Mandel, Naomi and Johnson, E. J. (2002): When Web Pages Influence Choice: Effects of Visual Primes on Experts and Novices. In Journal of Consumer Research, 29 (2) pp. 235-245

Martindale, Colin, Moore, Kathleen and Borkum, Jonathan (1990): Aesthetic Preference: Anomalous Findings for Berlyne's Psychobiological Theory. In The American Journal of Psychology, 103 (1)

Maslow, Abraham H. (1954): Motivation and Personality. HarperCollins Publishers

Maslow, Abraham Harold (1987): Motivation and Personality. HarperCollins Publishers

Moshagen, Morten and Thielsch, Meinald T. (2010): Facets of visual aesthetics. In International Journal of Human-Computer Studies, 68 (10) pp. 689-709

Moshagen, Morten, Musch, Jochen and Göritz, Anja S (2009): A blessing, not a curse: experimental evidence for beneficial effects of visual aesthetics on performance. In Ergonomics, 52 (10) pp. 1311-1320

Nake, Frieder (2005): Computer art: a personal recollection. In Proceedings of the 5th conference on Creativity cognition,

Nasar, Jack L. (1997): The Evaluative Image of the City. Sage Publications, Inc

Ngo, David Chek Ling, Teo, Lian Seng and Byrne, John G. (2003): Modelling interface aesthetics. In Information Sciences, 152 pp. 25-46

Nielsen, Jakob (1993): Usability Engineering. Boston, MA, Morgan Kaufmann

Norman, Donald A. (1988): The Psychology of Everyday Things. New York, Basic Books

Norman, Donald A. (1998): The Invisible Computer: Why Good Products Can Fail, the Personal Computer Is So Complex and Information Appliances Are the Solution. MIT Press

Norman, Donald A. (2004): Emotional Design: Why We Love (Or Hate) Everyday Things. Basic Books

Ortony, Andrew, Norman, Donald A. and Revelle, William (2005): The role of affect and proto-affect in effective functioning. In: Fellous, Jean-Marc and Arbib, Michael A. (eds.). "Who Needs Emotions?: The Brain Meets the Robot (Series in Affective Science)". Oxford University Press

Pandir, Muzeyyen and Knight, John (2006): Homepage aesthetics: The search for preference factors and the challenges of subjectivity. In Interacting with Computers, 18 (6) pp. 1351-1370

*Park, Su-e, Choi, Dongsung and Kim, Jinwoo (2004): Critical factors for the aesthetic fidelity of web pages: empirical studies with professional web designers and users. In Interacting with Computers, 16 (2) pp. 351-376*

*Park, Su-e, Choi, Dongsung and Kim, Jinwoo (2005): Visualizing E-Brand Personality: Exploratory Studies on Visual Attributes and E-Brand Personalities in Korea. In International Journal of Human-Computer Interaction, 19 (1) pp. 7-34*

*Parush, Avraham, Nadir, Ronen and Shtub, Avraham (1998): Evaluating the Layout of Graphical User Interface Screens: Validation of a Numerical Computerized Model. In International Journal of Human-Computer Interaction, 10 (4) pp. 343-360*

*Porteous, J. Douglas (1996): Environmental Aesthetics: Ideas, Politics and Planning. Routledge*

*Postrel, Virginia (2002): The Substance of Style: How the Rise of Aesthetic Value Is Remaking Commerce, Culture, and Consciousness. HarperCollins*

*Postrel, Virginia (2008). The Art of Healing - Magazine - The Atlantic. Retrieved 1 January 2011 from*

*Postrel, Virginia (2003): The Substance of Style: How the Rise of Aesthetic Value Is Remaking Commerce, Culture, and Consciousness. HarperCollins*

*Quinn, Jeffrey M. and Tran, Tuan Q. (2010): Attractive phones don't have to work better: independent effects of attractiveness, effectiveness, and efficiency on perceived usability. In: Proceedings of ACM CHI 2010 Conference on Human Factors in Computing Systems 2010. pp. 353-362*

*Rafaeli, Anat and Vilnai-Yavetz, Iris (2004): Emotion as a Connection of Physical Artifacts and Organizations. In Organization Science, 15 (6) pp. 671-686*

*Ravina, Enrichetta (2008). Love & Loans: The Effect of Beauty and Personal Characteristics in Credit Markets. Columbia Business School*

*Reber, Rolf, Schwarz, Norbert and Winkielman, Piotr (2004): Processing fluency and aesthetic pleasure: is beauty in the perceiver's processing experience?. In Personality and social psychology review an official journal of the Society for Personality and Social Psychology Inc, 8 (4) pp. 364-382*

*Reimann, M., Zaichkowsky, J., Neuhaus, C., Bender, T. and Weber, B. (2010): Aesthetic package design: A behavioral, neural, and psychological investigation. In Journal of Consumer Psychology, 20 (4) pp. 431-441*

*Rhodes, Gillian (2006): The evolutionary psychology of facial beauty. In Annual Review of Psychology, 57 (1) pp. 199-226*

*Santayana, George (1955): The Sense of Beauty: Being the Outline of Aesthetic Theory. Dover Publications*

*Sauer, Juergen and Sonderegger, Andreas (2009): The influence of prototype fidelity and aesthetics of design in usability tests: effects on user behaviour, subjective evaluation and emotion. In Applied Ergonomics, 40 (4)*

*Schaik, Paul Van and Ling, Jonathan (2009): The role of context in perceptions of the aesthetics of web pages over time. In International Journal of Human-Computer Studies, 67 (1) pp. 79-89*

*Schaik, Paul Van and Ling, Jonathan (2011): An integrated model of interaction experience for information retrieval in a Web-based encyclopaedia. In Interacting with Computers, 23 (1) pp. 18-32*

*Schenkman, Bo N. and Jonsson, Fredrik U. (2000): Aesthetics and Preferences of Web Pages. In Behaviour and Information Technology, 19 (5) pp. 367-377*

*Seligman, Martin E. P. and Csikszentmihalyi, Mihaly (2000): Positive Psychology: An Introduction. [Article]. In American Psychologist, 55 (1) pp. 5-14*

*Shaked, Gilboa, and Rafael, Anat (2003): Store environment, emotions and approach behaviour: applying environmental aesthetics to retailing. In The International Review of Retail, Distribution and Consumer Research, 13 (2) pp. 195-211*

*Simonson, Alex and Schmitt, Bernd H. (1997): Marketing Aesthetics: The Strategic Management of Brands, Identity and Image. Free Press*

*Sonderegger, Andreas and Sauer, Juergen (2010): The influence of design aesthetics in usability testing: effects on user performance and perceived usability. In Applied Ergonomics, 41 (3) pp. 403-410*

*Sun, Heshan and Zhang, Ping (2006): The role of affect in Information systems research. In: Zhang, Ping and Galletta, Dennis (eds.). "Human-Computer Interaction and Management Information Systems: Foundations (Advances in Management Information Systems)". M.E. Sharpe*

*Sutcliffe, Aiistair (2009): Designing for User Engagement: Aesthetic and Attractive User Interfaces. In Synthesis Lectures on HumanCentered Informatics, 2 (1) pp. 1-55*

*Sutcliffe, Alistair G. (2002): Assessing the Reliability of Heuristic Evaluation for Website Attractiveness and Usability. In: HICSS 2002 2002. p. 137*

*The Chromium Blog (2009). WebP, a new image format for the Web. Retrieved 1 December 2011 from The Chromium Blog: http://blog.chromium.org/2010/09/webp-new-image-format-for-web.html*

*Thuring, Manfred and Mahlke, Sascha (2007): Usability, aesthetics and emotions in human-technology interaction. In International Journal of Psychology, 42 (4) pp. 253-264*

*Townsend, Claudia and Shu, Suzanne B. (2010): When and how aesthetics influences financial decisions. In Journal of Consumer Psychology, 20 (4) pp. 452-458*

*Tractinsky, Noam (1997): Aesthetics and apparent usability: empirically assessing cultural and methodological issues. In: Proceedings of the SIGCHI conference on Human factors in computing systems 1997. pp. 115-122*

*Tractinsky, Noam (2004): Toward the study of aesthetics in information technology. In: Proceedings of the 25th International Conference on Information Systems December 12-15, 2004, Washington, DC, USA. pp. 771-780*

*Tractinsky, Noam (2006): Aesthetics in Information Technology: Motivation and Future Research Directions. In: Zhang, Ping and Galletta, Dennis (eds.). "Human-Computer Interaction and Management Information Systems: Foundations (Advances in Management Information Systems)". M.E. Sharpe*

*Tractinsky, Noam and Hassenzahl, Marc (2005): Arguing for Aesthetics in Human-Computer Interaction. In i-com Zeitschrift für interaktive und kooperative Medien, 4 (3) p. 66–68*

*Tractinsky, Noam and Meyer, Joachim (1999): Junkchart or Goldgraph? Effects of Presentation Objectives and Content Desirability on Information Presentation. In MIS Quarterly, 23 pp. 397-420*

*Tractinsky, Noam and Zmiri, Dror (2006): Exploring Attributes of Skins as Potential Antecedents of Emotion in HCI. In: Fishwick, Paul A. (ed.). "Aesthetic Computing". The MIT Press*

*Tractinsky, Noam, Katz, A. S. and Ikar, D. (2000): What is Beautiful is Usable. In Interacting with Computers, 13 (2) pp. 127-145*

*Tractinsky, Noam, Inbar, Ohad, Tsimhoni, Omer and Seder, Thomas (2011): Slow down, you move too fast: Examining animation aesthetics to promote eco-driving. In: 3rd International Conference on Automotive User Interfaces and Interactive Vehicular Applications AutomotiveUI 2011 November 30th-December 2nd, 2011, Salzburg, Austria.*

Tractinsky, Noam, Cokhavi, Avivit, Kirschenbaum, Moti and Sharfi, Tal (2006): Evaluating the consistency of immediate aesthetic perceptions of web pages. In International Journal of Human-Computer Studies, 64 (11) pp. 1071-1083

Tractinsky, Noam, Abdu, Rotem, Forlizzi, Jodi and Seder, Thomas (2011a): Towards personalisation of the driver environment: investigating responses to instrument cluster design. In International Journal of Vehicle Design, 55 (2) pp. 208-236

Tufte, Edward R. (1992): The Visual Display of Quantitative Information. Graphics Press

Turkle, Sherry (2005): The Second Self: Computers and the Human Spirit - Twentieth Anniversary Edition. The MIT Press

Udsen, Lars Erik and Jørgensen, Anker Helms (2005): The aesthetic turn: unravelling recent aesthetic approaches to human-computer interaction. In Digital Creativity, 16 (4) pp. 205-216

Ulrich, R. S. (1984): View through a window may influence recovery from surgery. In Science, 224 (4647) pp. 420-421

Venkatesh, Alladi and Meamber, Laurie (2008): The aesthetics of consumption and the consumer as an aesthetic subject. In Consumption Markets Culture, 11 (1) pp. 45-70

Veryzer, Robert W. and Hutchinson, J Wesley (1998): The Influence of Unity and Prototypicality on Aesthetic Responses to New Product Designs. In Journal of Consumer Research, 24 (4) pp. 374-385

Visser, Willemien (2009): Design: one, but in different forms. In Design Studies, 30 (3) pp. 187-223

Ware, Colin (2008): Visual Thinking: for Design. Morgan Kaufmann

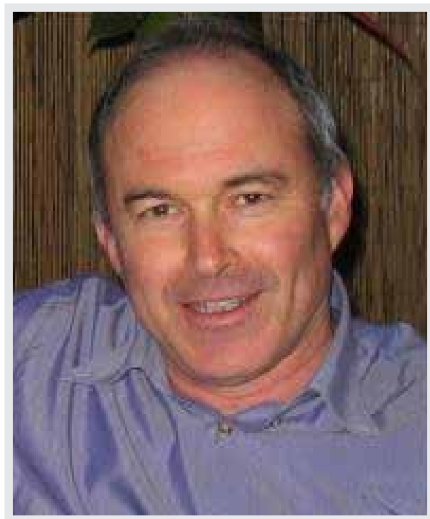
Whyte, Jennifer, Gann, D. and Salter, A. (2003): Design quality indicator as a tool for thinking. In Building Research Information, 31 (5) pp. 318-333

Winkielman, Piotr, Halberstadt, Jamin, Fazendeiro, Tedra and Catty, Steve (2006): Prototypes are attractive because they are easy on the mind. In Psychological Science, 17 (9) pp. 799-806

Winograd, Terry (1996): Bringing Design to Software. ACM Press

## ABOUT THE AUTHOR

### Noam Tractinsky



© ***Noam Tractinsky***

PERSONAL HOMEPAGE:
ise.bgu.ac.il/faculty/noam/noamt.html

I am an Associate Professor at the Department of Information Systems Engineering, Ben-Gurion University of the Negev, Beer Sheva, Israel. My main research interests revolve around the use of information technology. I have been involved in studies on IT project managers, online consumer behavior, the potential distraction of mobile technology, and aiding Alzheimer's patients. Currently I study various aspects of the interactive experience with an emphasis on aesthetics.

## Your notes and thoughts on chapter 19

Record your notes and thoughts on this chapter. If you want to share these thoughts with others online, go to the bottom of the page at: http://www.interaction-design.org/encyclopedia/visual_aesthetics.html

**NOTES:**

# Chapter 20

# Tactile Interaction

*by Ben Challis.*

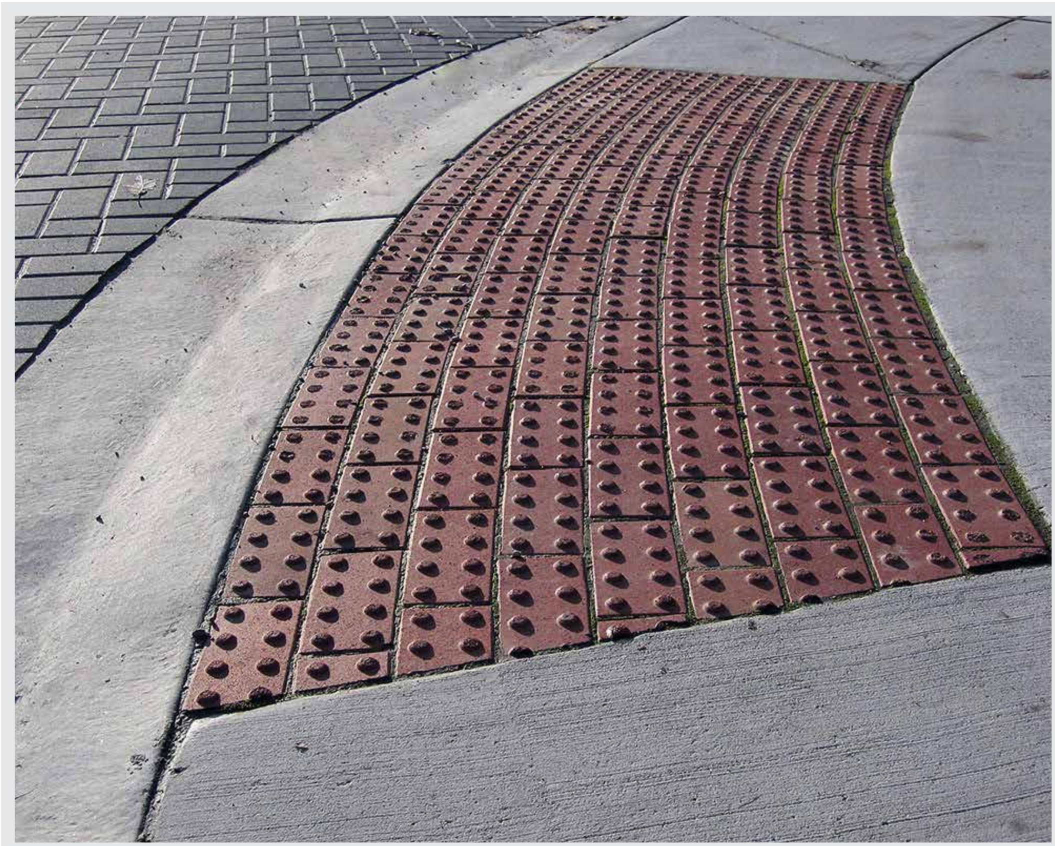
The following chapter describes a variety of ways in which Tactile Interaction may be used to enhance the human computer interface, i.e. the design of interactive products. Opening with a general discussion on a broad range of potential applications for Tactile Interaction, the chapter quickly moves onto to consider the key physical, perceptual and technological issues that are likely to influence the ways in which we can tap into this potentially rich source of interaction. A broad range of research topics are visited along the way with suggestions being offered for the fundamental design principles that should be considered within any interface that seeks to harness some level of tactile interaction.

## 20.1 INTRODUCTION

Let us consider why we might be interested in exploring tactile interaction within the design process at all? After all, is it not generally the visual channel that features most prominently within any given interface? In terms of display and feedback, this is often true; as users we perhaps expect to first *see* our available controls and options before we begin to interact with them, and similarly we probably expect to *see* the results of any actions we take. Interface design has long embraced the notion of offering additional methods of reinforcement where possible, with particular emphasis on the use of auditory feedback. So, any given action is likely to be reinforced through a combination of visual and auditory feedback, indeed, in an ever-expanding market for music and audio software, the feedback may be primarily auditory anyway; the music or sound itself. This undoubtedly powerful combination enables rich forms of both display and feedback, but the interaction that takes us from one to the other will generally be supported through *touch*.

The manner in which we select and manipulate these various physical and virtual objects will be through a combination of movement and touch or *haptic interaction*. This is perhaps where things begin to become interesting for though there is a considerable body of research and well documented ‘good’ practice on interface-design in terms of visual and auditory feedback, there is still very little on the specifics of ‘good’ design for haptic interaction. So, a first answer to our opening question could be that tactile interaction is perhaps being undervalued in terms of the potential source of feedback that it might offer. The primary focus has perhaps been too firmly fixed on the ‘doing’ rather than the ‘receiving’. So, in a world where visual displays are already quite cluttered, off-loading some of that information to the auditory channel could be of real benefit, but it might also be that some of this feedback is more immediately meaningful to our sense of touch. By way of example, the action of selecting and positioning a virtual on-screen fader might simply *feel* more meaningful than it can ever *sound*.

Figure 20.1: An example of tangible interaction design for “doing” — the “Reactable”. Note informational cues that inform the user of what each block does are visual, the tactility is how users take action.



*Courtesy of Richard Drdul. Copyright: CC-Att-SA-2 (Creative Commons Attribution-ShareAlike 2.0 Unported).*




Figure 20.2 A-B: Examples of tactile design for "receiving". These different textures feel different beneath the feet, and that change in texture informs blind (or texting) pedestrians when to stop or when to pay attention.

This argument alone is quite compelling, but let us now consider the numerous situations and environments in which a visually dominant display is either impractical or impossible. Perhaps the most immediately apparent example would be the design of non-visual interfaces for users who are blind. There are many issues to consider here, and some of these will surface later in this chapter. However, the key point for now is that the graphical user interface (GUI) will need to be translated and communicated effectively using non-visual means. There are some aids to draw upon here, screen-readers can use synthetic speech to read on-screen text or to describe a structural layout — but might it not be faster and more informing to display structural components in such a way that the user can 'feel' the controls? So, there are *extraordinary* needs for some computer users where a significant move away from a visually dominant interface is likely to be of benefit. However, there are all manner of occasions where the same is likely to be true for users who do not have additional individual needs. When controlling a vehicle, for example, or operating specialist machinery, it might simply be 'safer' to be able to maintain near-constant visual contact with our surroundings. Indeed, the location of controls for in-car entertainment systems is now often found on the steering column — within easy reach and with no requirement to look away from the road. In some situations, the immediate environment may not offer enough light to easily see what is happening. At the extreme end of these scenarios, perhaps an environment is currently under emergency lighting, and at the less safety-critical end, perhaps the living room is a little too dark to easily control the home-cinema by remote control.

So, there is a broad range of reasons why, as designers, we might wish to explore non-visual modes of communication. Our interest within the context of this chapter, though, is with tactile interaction, and later we will consider the various technologies (existing and emerging) that can be drawn upon to make this happen. However, as with all other forms of interaction and feedback there are physical and perceptual limits and boundaries that will influence how effective any new design might be.

## 20.2 THE PSYCHOLOGY OF TOUCH

To begin to appreciate how tactual interaction might be successfully integrated into the human-computer interface, it is essential to understand how the human body retrieves and processes information about its immediate surroundings. This happens at two levels: physical and perceptual. At the physical level, our peripheral nervous systems gather information using a number of different nerve types each of which is sensitive to a particular type of stimulus. All the information gathered by the peripheral nervous system is conveyed through the central nervous system to its ultimate focal point: the brain. It is here that the information is interpreted and then acted upon, and it is this process of interpretation that constitutes the perceptual level. For the purposes of this chapter, we shall avoid any further discussion on the physiological make-up of the human nervous system, but for anyone considering exploring tactile interaction within a design context, this really should feature as part of your recommended further reading.

Loomis and Lederman (1986) provide a useful overview of the three aspects of interpreting information through touch that could be referred to collectively as *tactual perception*. They state that there are two fundamental and distinct senses that together provide us with a sense of touch: the cutaneous sense and kinesthesis. The cutaneous sense provides an awareness of the stimulation of the receptors within the skin, whereas the kinesthetic sense provides an awareness of the relative positioning of the body (head, torso, limbs etc.). Perception that involves one or more of these can be regarded as *tactual perception* and there are, therefore, three forms of such perception.

**Tactile perception** is solely dependent upon variations in cutaneous stimulation by such actions as tracing a pattern upon an individual's skin. Tactile perception alone means that the individual in question must be static; otherwise the kinesthetic sense will be incorporated.

**Kinesthetic perception** is concerned with variations in kinesthetic stimulation. However, tactual perception without contribution from the cutaneous sense can only really be achieved under contrived circumstances such as using anaesthetic to suppress the cutaneous contribution.

**Haptic perception** is the form of tactual perception that involves both tactile and kinesthetic perception, and it is this that we use on an everyday basis to explore and understand our surroundings using touch.

The level of control that an individual has at any given time over the collection of information by either the cutaneous or kinesthetic sense leads to the following five tactual modes, where for the first three modes there is no control. Indeed, it can be seen from these definitions that of the five modes only the last, *active haptic perception*, is likely to be of real significance within the design of interfaces that employ tactile interaction.

1. Tactile perception - Cutaneous information alone.
2. Passive kinesthetic perception - Afferent kinesthesis.
3. Passive haptic perception - Cutaneous information and afferent kinesthesis.
4. Active kinesthetic perception - Afferent kinesthesis and efference copy.
5. Active haptic perception - Cutaneous information, afferent kinesthesis and efference copy.

*Courtesy of Amanda M. Williams. Copyright: CC-Att-SA-3 (Creative Commons Attribution-ShareAlike 3.0).*

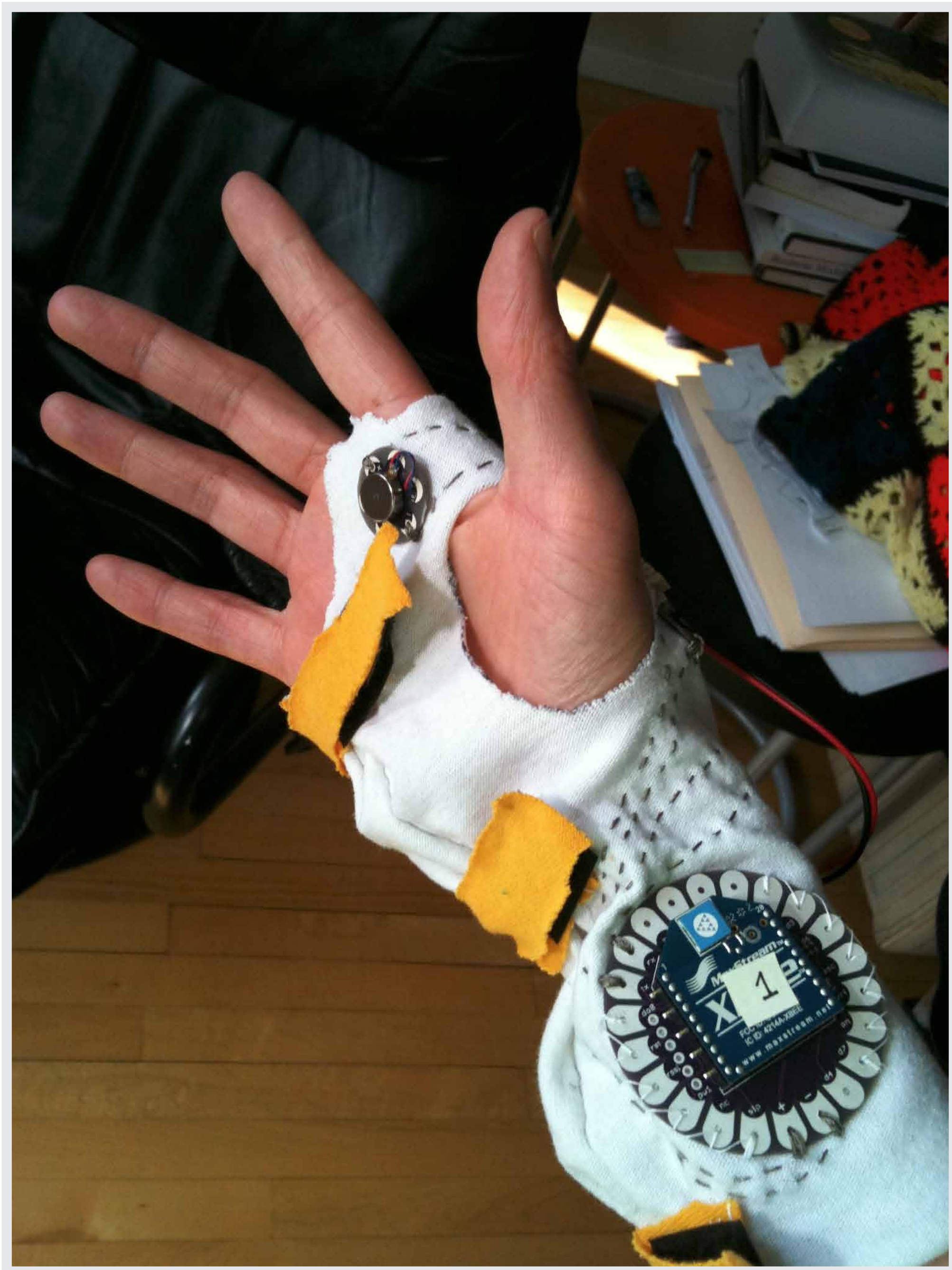





FIGURE 20.3 A-B: Vibrotactile feedback, such as this prototype glove uses, takes advantage of the user's cutaneous sense.

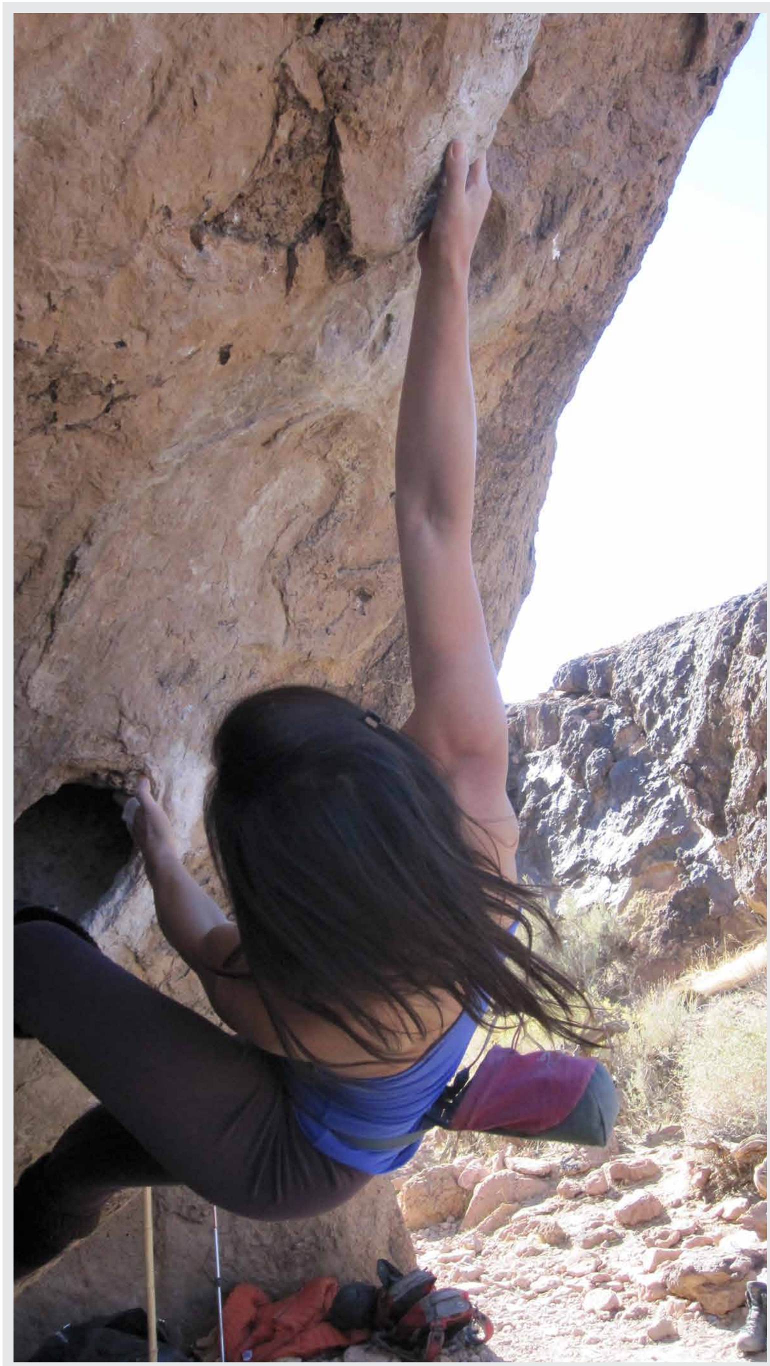


Figure 20.4: The kinesthetic sense provides an awareness of the relative positioning of the body.

## 20.3 THE PRACTICALITIES OF TACTILE INTERACTION

Ultimately, the design of any human-computer interface that incorporates tactile interaction is bound to share many of the same considerations over construction that need to be made when designing, for example, tactile diagrams, and there is a considerable body of research within this area that we can turn to. Before we consider some of these perceptual properties in depth, let us first quickly consider the relationship between *touch* and *sight*.

### 20.3.1 Vision versus Touch

Our general reliance on vision suggests that if a conflict were to arise between our sense of touch and our sense of sight, it would be the visual aspect that becomes dominant. Such conflict of sensory information formed the basis for research by Rock and Victor (1964) who used an optical cylinder to give a subject group the visual impression that a solid square object was really rectangular. To the subjects in the experiment, the tactual impression was unchanged. Subjects who were presented with such conflicts generally made similar judgements on size and shape as those members of a control group who were only presented with visual information.

However, McDonnell and Duffett (1972) suggest that Rock and Victor's experiment may have had methodological failings that could have produced a bias in favour of vision. In their reworking of the original experiment, subjects were asked to examine blocks of wood on the top surface of a table. To examine the lower part of the blocks, the subjects had to feel beneath the table. The impression that the subjects were given was that the blocks went through the table, whereas there were really two blocks being used to give that impression. In effect, each block that the subjects examined was a pair of blocks where the width of the wood and the protrusion on either side of the table was the same but the lengths could be different. Five pairs of blocks were used with the discrepancy ratios of 1:1, 1.29:1, 1.67:1, 2.2:1 and 3:1. After examining each block, the subjects would have to choose the best match from a group of comparison blocks that had no length discrepancy. The mean scores for the group showed that the subjects had chosen a comparison block that was a compromise between their visual and tactual impression. However, closer examination showed that the subjects had made their choice conform to either the visual or tactual impression and that there was considerable diversity of response pattern such that visual dominance could not be regarded as significant.

Heller (1992) has also shown that Rock and Victor's original findings should not have been applied so generally as to state that vision will always be the more dominant sense when there is conflict between touch and vision. In Heller's experiment, subjects were required to explore the embossed letters *p, q, b, d, w* and *m* whilst looking at them in a mirror. A subject would, for example, be exploring the letter *w* whilst looking at the letter *m*. When asked to identify the letters, there were a wide range of responses where the majority relied on touch, some on a compromise between the two and only one relied on vision.

It seems, therefore, that visual dominance versus tactual dominance should not be thought of as a dichotomy as there is evidence of compromise between the two senses when they are in conflict. In addition, this level of compromise is likely to be highly individual and will also be affected by a bias towards the suitability of one or both senses to the nature of the task.

### 20.3.2 Visual to Tactile Mapping

Klatzky and Lederman (1987) argues that many tactile diagrams will probably have problems due to the limiting nature of spatial resolution describing, for example, line drawings that are seen well might not be felt that well because the scale simply becomes too small for such limited bandwidth. Klatzky and Lederman also suggest that a fundamentally flawed model of haptic processing is often employed in the design of tactile graphics displays. Referred to as the *image mediation model*, the "hand functions like a roving eye that is badly in need of glasses" (Klatzky and Lederman 1987). The assumption made is that a spatial image is produced which is equivalent to one produced using vision. However, this image is affected by factors such as the low resolution of the haptic sensors along with demand placed upon memory due to the nature of exploring an 'image' over time. This image is then passed to the visual system's image interpreters resulting in a mental-image which perhaps seems like the original diagram was examined visually by someone with poor eyesight.

In contrast to this model, Klatzky and Lederman suggest that the haptic system has its own perceptual system and interpretive processors. Haptics and vision are simply different ways of perceiving, although at a much higher cognitive level there may be some convergence of the haptic system into the visual. Perhaps the most significant suggestion from Klatzky and Lederman is that the haptic system is not an efficient mediator of images. By way of example, it is described that if an individual is asked to think of looking at a cat, he/she is likely to visualise the shape of its body and perhaps the colour and pattern of its fur. However, if the same person is then asked to think of touching a cat, a very different image will result; the softness of its fur and perhaps the warmth of its body.

During studies by Klatzky and Lederman into exploration of three-dimensional objects, a series of methods for exploration were recorded. Of particular interest was that substance dimensions (e.g. hardness and texture) can be extracted quickly and reliably, whereas structural information is extracted slowly and is error prone. If it is assumed that the haptic system will favour encoding mechanisms which produce maximum return for minimum effort, then it would be expected that an economical system would favour substance based exploration and encoding. One particular study involved subjects sorting objects that varied along four dimensions: hardness, surface-roughness, size and shape. There were three variations for each dimension, and all possible combinations were covered within the whole group. The subjects were simply asked to place objects that were similar into a common bin. By examining the objects within each bin, the dimension that had proved most salient could be established. For example, had all the rough objects been in one bin, the medium rough in another and the smooth in another, then surface texture would have been the most salient dimension.

The results varied depending on the exploratory conditions which were allowed such that when denied sight, the subjects would prefer the substance dimensions of either hardness or texture. This pattern was unchanged when similarity was defined as how objects 'felt', but there was a large shift towards reliance on structural dimensions when similarity was defined in terms of objects being similar 'visually'. Finally, when the subjects were allowed to see the objects, the structural dimensions were the strongest again. Although these studies were carried out using three-dimensional objects, the findings might be just as valid within tactile diagram design although only contour and texture would apply. In particular, if a display is to be partly visual but also reinforced with some tactile elements, Klatzky and Lederman's findings might help establish how best to distribute the information between the visual and haptic channels.

### 20.3.3 Line Symbols

Perhaps the simplest tactile object is that which represents the drawing primitive of a line. Straight and curved lines could be considered as the building blocks of most graphical representations, and graphs, maps, and other diagrams can rely heavily on the use of lines within their construction. There are two issues that quickly become apparent when using raised lines as tactile substitutes for visual ones:

1. How easily can they be traced?
2. How easily can different line widths be distinguished from one another?

On the traceability of lines, a comparison study has been made by Bentzen and Peck (1979) which addressed the issue of which styles of lines are easiest to follow by tracing with a finger. Four line styles were used: single continuous (smooth), double continuous, single dotted (rough) and double dotted. These were chosen because they seem to be accepted as the four most commonly used, but without any supporting evidence as to which is superior. Other than identifying which lines are generally the easiest to trace, Bentzen was also interested in how tracing can be affected within two particular scenarios:

- Displays in which lines do not have intersections.
- Displays in which lines do have intersections. Rough against smooth and single against double.

Two displays were created using embossed plastic sheets. A simple display used all four line types, without intersection, which all included a right angle, an obtuse angle and an acute angle and a semicircle of 1.5in. These were connected using three straight-line sections of 3in and three of 1.5in. The complex display had the four line styles using the same tracing components, but each line intersected the other three at some point.

The conclusions reached were that the performance of rough against smooth lines is not significantly different and is therefore not a real design issue in the use of tactile diagrams. Single lines, rough or smooth, are preferable to double lines (0.25in apart) in tactile displays that do not have intersecting lines. No real conclusion was found with double over single lines, other than double did perform better than single at intersections. This, however, was believed to be particular to the design of the displays used and was therefore not significant at this stage. One final observation was that a single narrow line intersected by a double is an undesirable feature.

An experiment by Lederman and Campbell (1983) explored the use of raised lines with tangible graph displays for blind people. Four different methods of presenting the graphs were used.

- **No-grid** - besides the tics on the major axes, there were no grid lines to be found in the main graph-area.
- **Grid-on-graph** - the tics extended across the main area in a grid format.
- **Grid-on-overlay** - the grid was an overlay to the no grid version. The subject could keep dropping the overlay down onto the graph.
- **Grid-on-underlay** - the opposite to grid-on-overlay. Here, the grid was the underlay part, and the graph could be dropped down onto the grid.

Three line styles were used: smooth, large-dotted and small-dotted. From a traceability perspective, the results were encouraging, but with graphs where all three line-types were present in close proximity, a significantly longer time was taken to perform tasks and these were more likely to be inaccurate. Some notes by the investigators mention that when initially presented with a graph, a subject would tend to explore its general format such as major axes, symbols and labels. Sandpaper squares at the corners were used to great effect in that by checking the squares and major axes, a broad sweep of the hand allowed the subject to become swiftly familiar with the graph's dimensions and proportions. Individual differences were observed as some people would use both hands to explore, whilst others would prefer to keep one hand on the origin as a point of reference. It also appeared that both the no-grid and grid-on-graph formats were preferred to the other options; both seemed to be equally easy to use.

One other exploration of the factors that affect the discriminability of tactile lines concerned the perception of one line as being wider than another (Berla and Murr 1975). In essence, five standard line thicknesses were each compared with a set of lines half of which were narrower and the rest thicker. The five standard line thicknesses used were 0.1, 0.15, 0.2, 0.25 and 0.3cm wide. The 0.1cm line had six progressively narrower lines and six wider, in steps of 0.01cm, whereas all the other lines used steps of 0.013cm. Each line was approximately 0.64cm in height. When presented with one standard line and one variable, the subject had to indicate which was widest. The results showed that the percentage of extra width that was required for an individual to perceive a line as wider than the standard decreased as the standard increased. For example, the thinnest standard line (0.1cm) needed to be 20% to 48% wider, whereas the thickest line (0.3cm) needed to be 11% to 27% thicker. However, this trend did not continue without finish. When the standard lines included thicknesses of 0.64cm and 1.27cm, performance began to diminish. This was attributed to the line width beginning to extend beyond the width of an average fingertip with the result being that both sides (or edges) to the line could not easily be perceived. As the results stand, though, there is a very useful set of highly distinguishable line widths that can safely be employed where difference in line width will play a significant part in a tactile display.

Nolan and Morris (1971) suggested that only around eight tactile linear symbols could be in use at any given time before similarities would occur and that there was likely to only be a maximum of ten. A study by James and Gill (1975) showed there to be ten such distinguishable linear symbols, but they were unable to progress beyond Nolan and Morris's upper limit.

### 20.3.4 Point Symbols

Point symbols are probably best described as those symbols which are designed to be explored with only minimal movement of the fingertip. An important aspect of the use of such symbols is how well they can be perceived in contrast to the background. This is commonly referred to as the *figure-ground problem*. One aspect of this is whether raised symbols are easier to recognise than incised. An assumption that there is no particular difference between either has been tested (Nolan and Morris 1971) and in contrast to popular belief it seems that there *is* a significant difference. Braille readers were asked to trace a tactile symbol and then find that same symbol from a set of five possibilities. This was carried out using a set of raised figures and an equivalent set of incised figures. Nolan's result showed that, in fact, raised figures are significantly superior to incised figures. Although there was an increase in error rate with the incised symbols, the significant result was that there was a required increase in reading time of around 38%. This implies that most tactile symbols will be read faster and more accurately if raised figures are used.

Research by Lambert and Lederman (1989) into the relative legibility and meaningfulness of tactile map symbols has shown that there are three categories of tactile symbol to be used. Some symbols have inherent meanings such as a telephone shape, some can imply a meaning such as a pointed symbol for stop and some can have quite arbitrary meanings such as using a square to represent a washroom. The size of the symbols used was between 0.635cm and 1.27cm per side. These were based on previous observations and studies with blind people. The notable feature that Lederman discusses is that, although some people will instantly prefer or recognise certain symbols, the question of whether a symbol possesses inherent meaning is not paramount. With prolonged use, any symbol will begin to imply a specific meaning and to that effect there will be a learning overhead expected in the successful implementation of any tactile diagram.

### 20.3.5 Areal Symbols

The term *areal symbol* is used to describe those areas of a tactile diagram which use either a texture or a tactile pattern to communicate information. The term tactile-pattern perhaps implies an area covered with incised or raised symbols which can easily be perceived as being identifiable and distinguishable from similar patterns. However, Lederman (1982) makes the point that a fundamental aspect of tactile pattern perception is the perception of texture. She uses examples of the 'smoothness' of a baby's skin, 'roughness' of sandpaper, 'softness' of cashmere, 'rubberiness' of elastic and the 'slipperiness' of ice. Texture alone provides much of the tactual feedback that we need to decide whether one areal tactile pattern is highly distinguishable from other areal patterns.

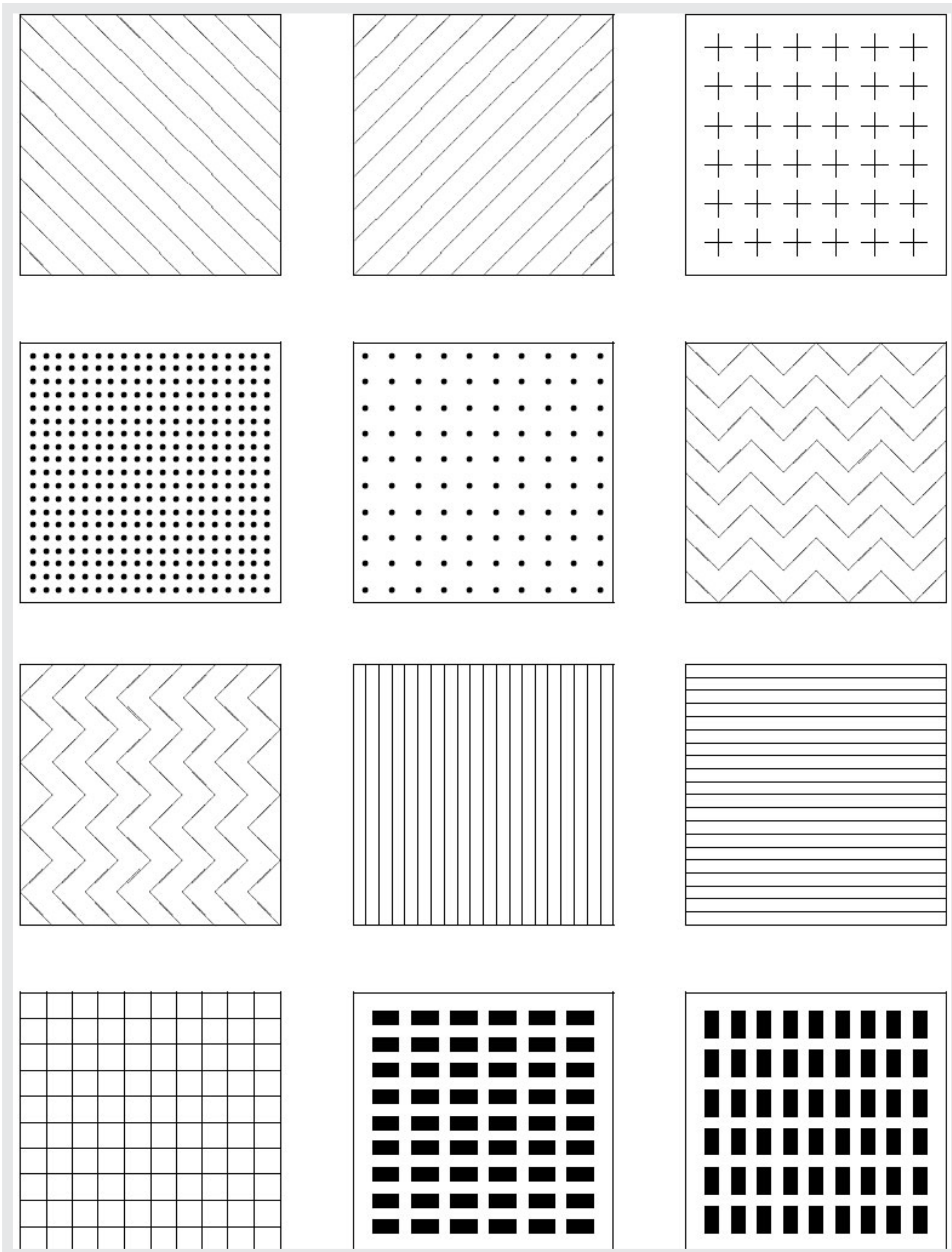


Figure 20.5: Some examples of possible areal tactile symbols.

Lederman discusses some of the known issues with the perception of texture outlining how she has studied the perception of roughness using aluminium plates with linear grooves cut into the surface. Her results have shown that the main factor affecting the perception of roughness is the width of the groove cuts in relation to their spacing. However, she has also shown that applied force is the next most significant factor, followed by hand speed. A greater applied force produces results that show a higher ratio of perceived roughness, and this can be shown to be true with a gradually decreased hand speed. These findings are significant in that they could play an important part in the design of appropriate tactile patterns to employ as areal symbols.

Loomis (1981) describes a series of limitations that affect tactile pattern perception: spatial resolution, interactions between stimuli more widely spaced than the resolution limit (interference), temporal resolution, perceptual integration and limited attention.

Spatial resolution - This is normally associated with the two-point limen test where the two points of a draftsman's compass are placed onto an individual's skin in close proximity to each other. The distance of interest is the threshold at which the two points are perceived as one. However, Loomis breaks spatial resolution into three further factors. Firstly, there is the mechanical property of the skin. When a point is placed on the skin, the gradient of skin deformation will be considerably less than that of the stimulus. Secondly, there is the property of mechanical wave spreading. When a point is stimulated, travelling waves are produced which, when picked up by other mechanoreceptors, result in a 'blurring' effect. Lastly, there is the neural organisation that means that spatial resolution is dependent upon:

1. The density of mechanoreceptive units in a particular area of the skin.
2. The size and sensitivity of these units.
3. The number of neurons in the cortical projection areas that represent that same field.

**Interactions between widely spaced stimuli** - The two-point limen test is a clear example of this, but the focus is to assess the distance at which two stimuli are perceived as one. Having two such stimuli presented to body locations that are wide apart, such as separate hands, can lead to a phantom sensation at a point between the two. Another example of phantom sensations occurs when rapid stimuli are presented to one point and then immediately to another. The resulting feeling is of an evenly spaced series of sensations travelling from one location to the other. Cutaneous masking can also occur in that one stimulus can be completely masked by the presence of another much stronger stimulus.

**Temporal resolution** - The minimum period with which an individual can still clearly perceive two brief pulses as being separate. Loomis states that a number of experiments into this have arrived at estimates ranging from 2ms to 40ms.

**Perceptual integration** - The way in which all this information is used at a cortical stage even if it were to arrive there with no significant loss of detail. The suggestion is that, unlike with visual perception, the information from a stimulus pattern may fail to be recognised.

**Limited attention** - Even if the information was perceived without loss and integrated cortically, the individual might still not perceive the pattern as a result of insufficient attentional capacity. With the visual channel, it is widely acknowledged that people possess the ability to focus their attention when the quantity of information begins to exceed their processing capacity. The tactual channel does not seem as efficient for this form of focused attention.

When it comes to the practical application of tactile patterns as part of a display, Lederman and Kinch (1979) have provided a review of existing work in the field. A general conclusion that can be reached is that although there are around forty tactile patterns that can be found that are easy to recognise only up to eight can be used together. It is very difficult to find any more than eight patterns that can be used as a group without ambiguities creeping in. As an example, a typical poor choice of patterns would be the use of diagonal lines in opposing directions to signify separate functions. So, within the example set of twelve areal symbols shown earlier in Figure 20.5, there are already ambiguities beginning to creep in such that rapid distinction between certain pairs of symbols will be inhibited. A significant effect that Lederman discusses is the use of height to provide a filtering method within a tactile display. Symbols could be presented at one of three heights to provide an indication of significance, and through using a sweeping hand movement, this could be used as way of filtering out unwanted information.

Figure 20.6: A tactile overlay for conveying the structural information from a page of music notation. Made from vacuum-formed pvc, the overlay displays key elements such as bars, lines to a page, repeat marks etc. Similar overlays were used within an experimental system for providing access to music notation for blind people (Challis and Edwards 2000). Interacting with the overlay would allow the user to extract speech and audio descriptions of rhythmic and melodic content.





FIGURE 20.7 A-B: Examples of other overlays from the 'Weasel' non-visual music notation system. Here, each overlay is housed within an Intellikeys touch-sensitive tablet; a device used quite commonly in special needs education.

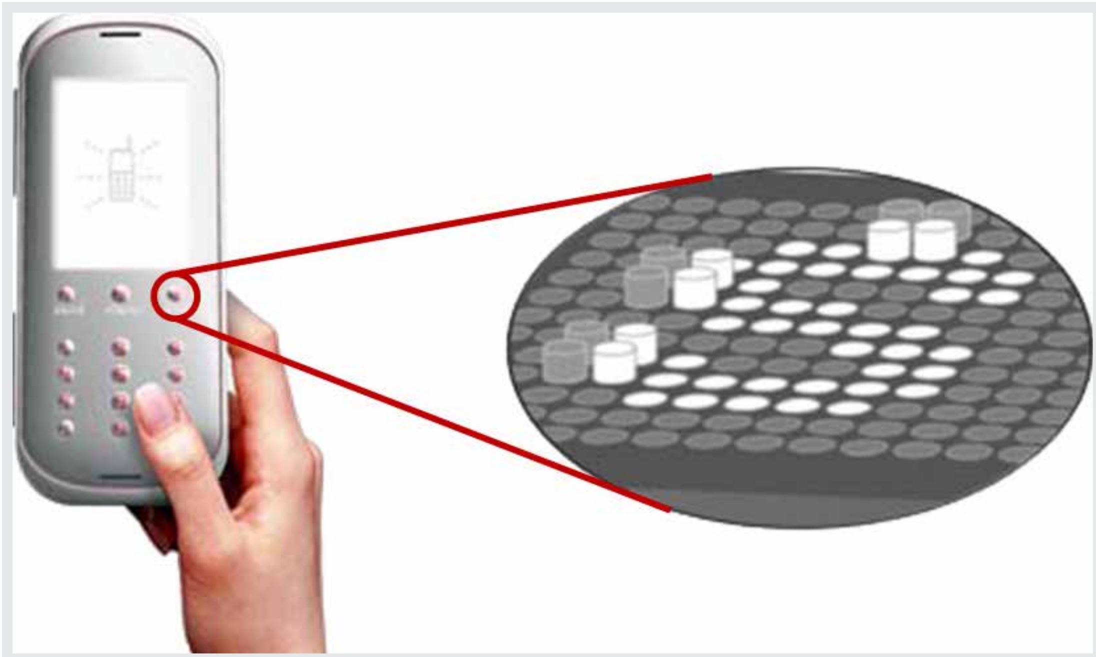


Figure 20.8: Taxels can be thought of as being tactile equivalents to pixels. The illustration here suggests how an array of taxels might be used to create a dynamic interface for a mobile phone.



### 20.3.6 Strategies for Exploration

Berla (1972) defines three problem areas in the use of tactile diagrams by blind people: *legibility*, *organisation* and *strategies for exploration*. Of these three problems, the last is of particular interest as it is concerned with how individuals adopt different methods for exploring diagrams. The results of an experiment in this area by Berla showed that some individuals will use a one-handed scanning strategy and others would use two. Those who used two hands proved to maintain a superior sense of orientation and position within the diagram than those who used only one hand. He suggests that this is the result of being able to use one hand as a reference point. The strategies that Berla identified can be described as:

- **Horizontal-unidirectional** - where a hand is moved horizontally across the page and is returned to the beginning of the line before progressing down or up the page for the next scan.
- **Horizontal-bidirectional** - where a hand traces in one direction across the page, is lowered or raised to the next line of scanning and then scans back across the page.
- **Asymmetrical horizontal scan** - where both hands are placed at the centre of the diagram and then moved outwards in opposite directions and then brought back to the centre again. The hands are then lowered or raised to the next scan line.
- **Vertical-unidirectional** - where a hand traces vertically across the page and is returned to the beginning of the line before progressing right or left for the next line of scanning.
- **Vertical-bidirectional** - where a hand traces vertically across the page, is moved left or right to the next line of scanning and then scans back vertically across the page.
- **Perimeter or 'clock-face' scan** - where a hand traces the full perimeter of the diagram and then is moved successively closer to the middle, scanning the smaller, inner perimeters.
- **Bounded search** - where arbitrarily sized 'boxes' are superimposed onto the diagram by the reader. Searches will be limited to particular boxes.
- **Density distribution scan** - where a hand is used to swiftly determine in which areas the majority of symbols can be found. The least populated areas are then explored first.
- **Spoked wheel scan** — where one hand is used as a reference point for the other hand which scans from the centre outwards, gradually moving around the diagram.

Berla concludes that these strategies all have relative merits and disadvantages which, therefore, make it difficult to suggest the 'ideal' strategy to adopt. He also describes how the most appropriate approach might be to teach all of these strategies to people who will be working with such diagrams. They will then be able to apply the most suitable strategy to any given exploration task. Berla and Butterfield (1977) has also suggested that individuals will probably need to be trained how to use a particular type of tactile diagram before any significant level of success can be attained. In his studies, he showed that a student's performance at distinguishing between, and understanding, tactile symbols increased if the student was first trained in line tracing and distinctive features analysis. There can be, perhaps, an expectancy for an individual to simply understand such tactile diagrams without any prior experience. This is an area that Berla underlines as being important when intending to design and implement a tactile display.

### 20.3.7 Braille Symbols

The Braille system is a method widely used by people who are visually impaired to read and write. Braille was devised in 1825 by Louis Braille, a blind Frenchman. Each Braille character, or cell, is made up of six dot positions, arranged in a rectangle containing two columns of three dots each.

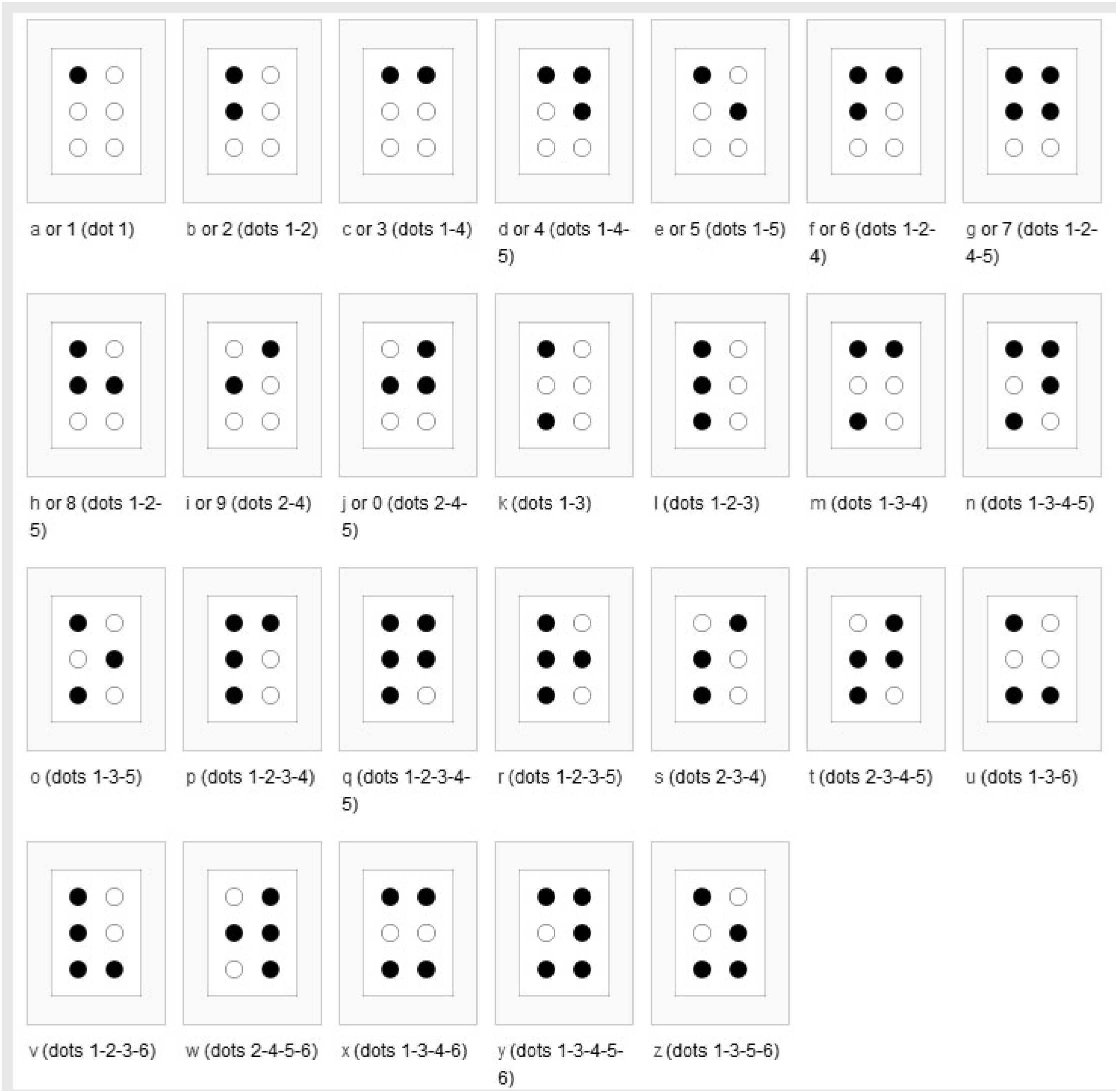



FIGURE 20.9: The Braille symbols.



Although braille symbols are not an integral aspect of tactile diagram design, each cell is effectively a tactile symbol, and factors that affect the legibility and perception of such cells could therefore be of interest. One particular piece of research into braille and pattern perception studied whether braille cells are perceived as groups of individual dots or as outline shapes. A series of experiments by Millar (1985) used partial outlines to represent the letters that would normally be shown in braille. These were not tactile outlines of the ordinary printed letters; instead they were joined up versions of the braille cells. Millar's results showed that braille letters can be read and recognised significantly faster with cells of dots rather than joined lines. This implies that small tactile symbols might also benefit from being made of patterns of dots rather than outlines.

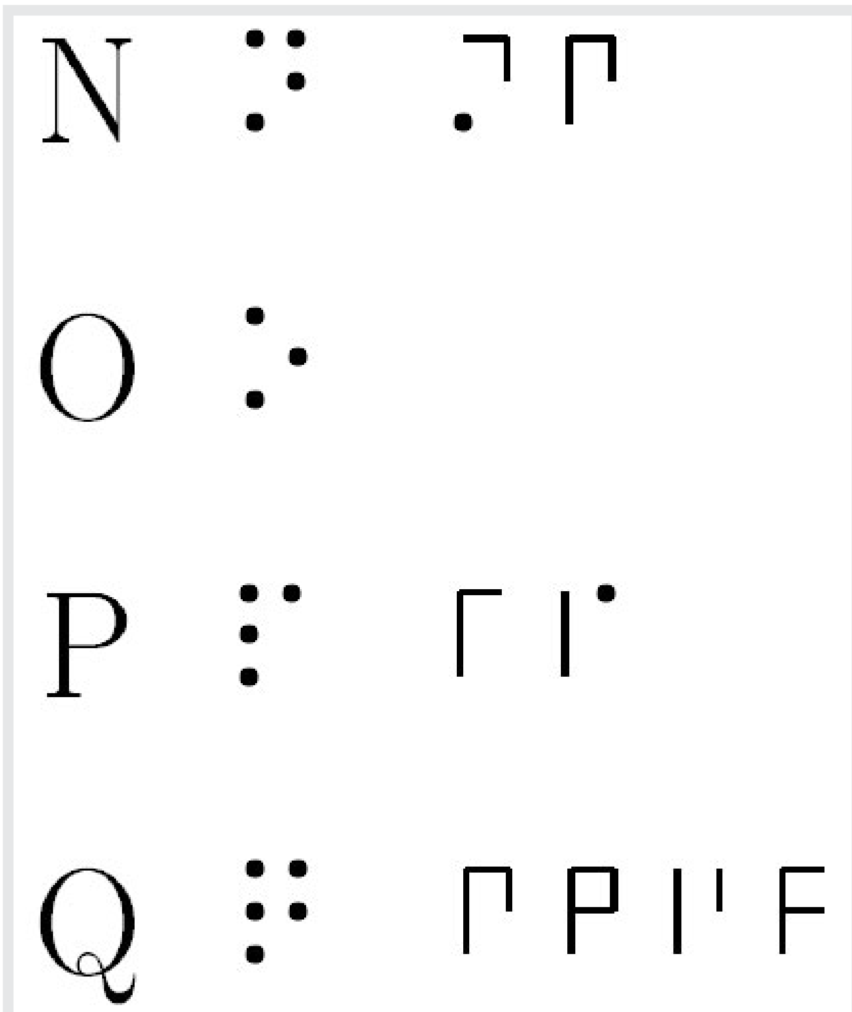



Figure 20.10: Examples of the outline shapes used to replace braille cells by Millar (1985).

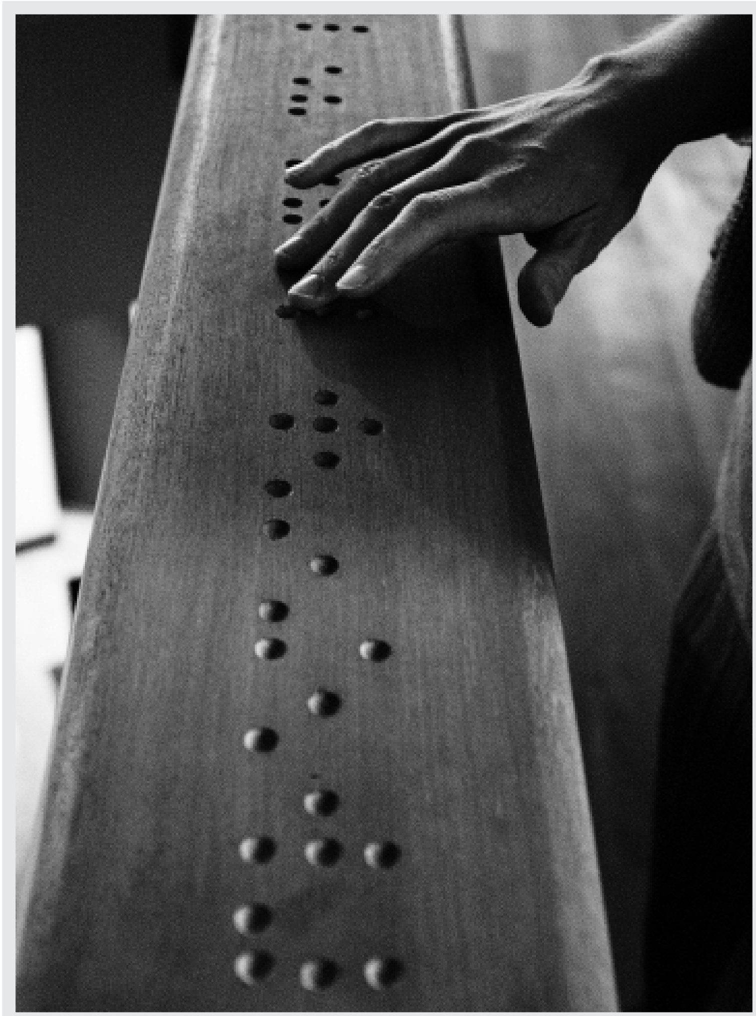


FIGURE 20.11: Wood-carved braille code of the word ‘premier’ (French for “first”).

COURTESY OF CHRISTOPHE MOUSTIER. COPYRIGHT: pd (Public Domain (information that is common property and contains no original authorship)).

## 20.3.8 Design Principles for Tactile Interaction

A preliminary set of design principles for including tactile information within the human-computer interface has been proposed (Challis and Edwards 2000, Challis and Edwards 2000). Using music notation as an example of a complex graphical information type, an experimental system was developed that could enable blind music-learners to adopt a non-visual multi-modal approach to reading music notation. Static overlays were used in conjunction with a resistive touch-pad to create interactive pages where a user could 'feel' the structure of the page layout and then select appropriate levels of information retrieval using audio and synthetic speech. Three founding principles were adopted at the outset of the study covering *consistency of mapping*, the *use of height* and the *use of static data* with additional principles being identified within the study to address aspects such as *size of display*, *visual-to-tactile mapping*, *simplicity of symbol design* and perhaps most significantly *empty space*; this simply introduces areas where the user has no information on where or how to explore. One simple but key observation put forward was that tactile diagrams might not *look* that good and that the design is not likely to benefit from over-reliance on direct visual-to-tactile mapping.

## 20.4 TACTILE INTERACTION IN THE HUMAN COMPUTER INTERFACE

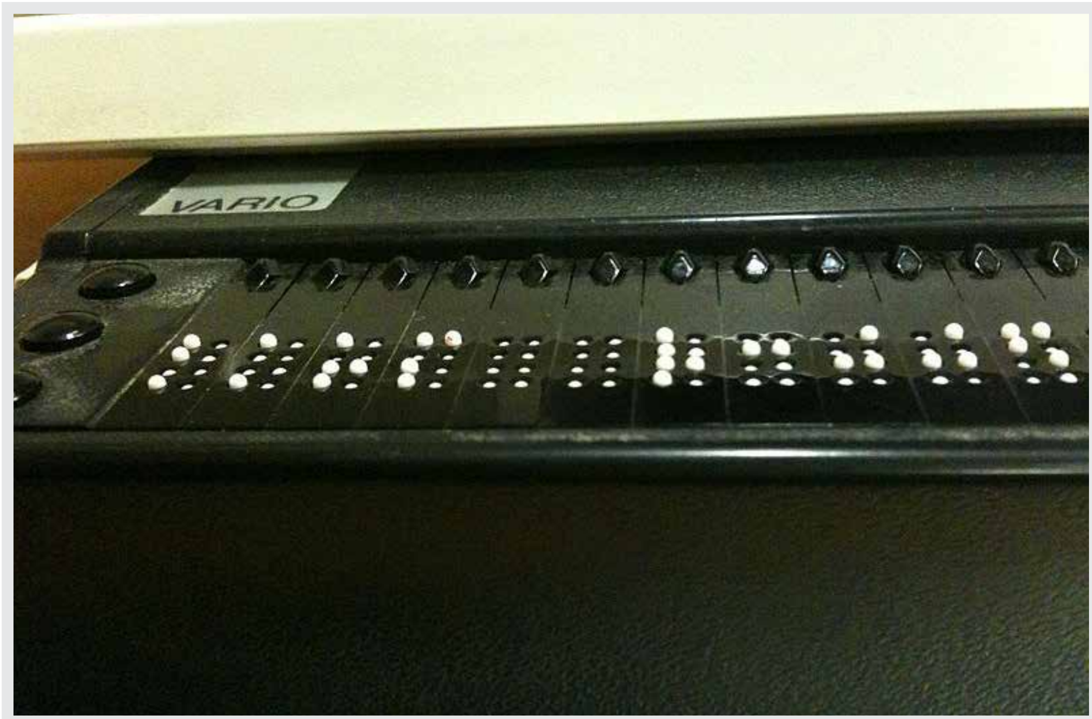
When employing technologies for harnessing tactile interaction, there are three broad approaches that might be considered: static tactile displays, dynamic tactile displays and force-feedback technology. There are clear merits and constraints associated with each such that the nature of interaction task will dictate which category will be most effective. Indeed, Oakley et al. (2000) have suggested definitions by which such different technologies can be categorised, this being based on the sensory system that is most affected by the interaction.

- Haptic - Relating to the sense of touch.
- Proprioceptive - Relating to the sensory information about the state of the body (including cutaneous, kinesthetic and vestibular sensations.)
- Vestibular - Pertaining to the perception of head position, acceleration and deceleration.

- Kinesthetic - Meaning the feeling of motion. Relating to sensations originating in muscles, tendons and joints.
- Cutaneous - Pertaining to the skin itself or the skin as a sense organ. Includes sensation of pressure, temperature and pain.
- Tactile - Pertaining to the cutaneous sense, but more specifically the sensation of pressure rather than temperature or pain.
- Force Feedback - Relating to the mechanical production of information sensed by the human kinesthetic system.

### 20.4.1 Static and Dynamic Tactile Displays

Where a permanent display is appropriate, an interactive information display, for example, or the controls on a dedicated device, then a static overlay superimposed upon an appropriate touch-responsive surface (e.g. resistive, capacitive, infrared, acoustic wave or force-sensing) will be practical whilst offering fine level of detail if required. In contrast, a dynamic display can offer increased flexibility as the interface is not 'tied' to a particular layout. Refreshable braille displays go some way towards offering this flexibility, but the size of the technology currently available does not lend itself to the creation of tactile symbols of a useful resolution.




*Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.*

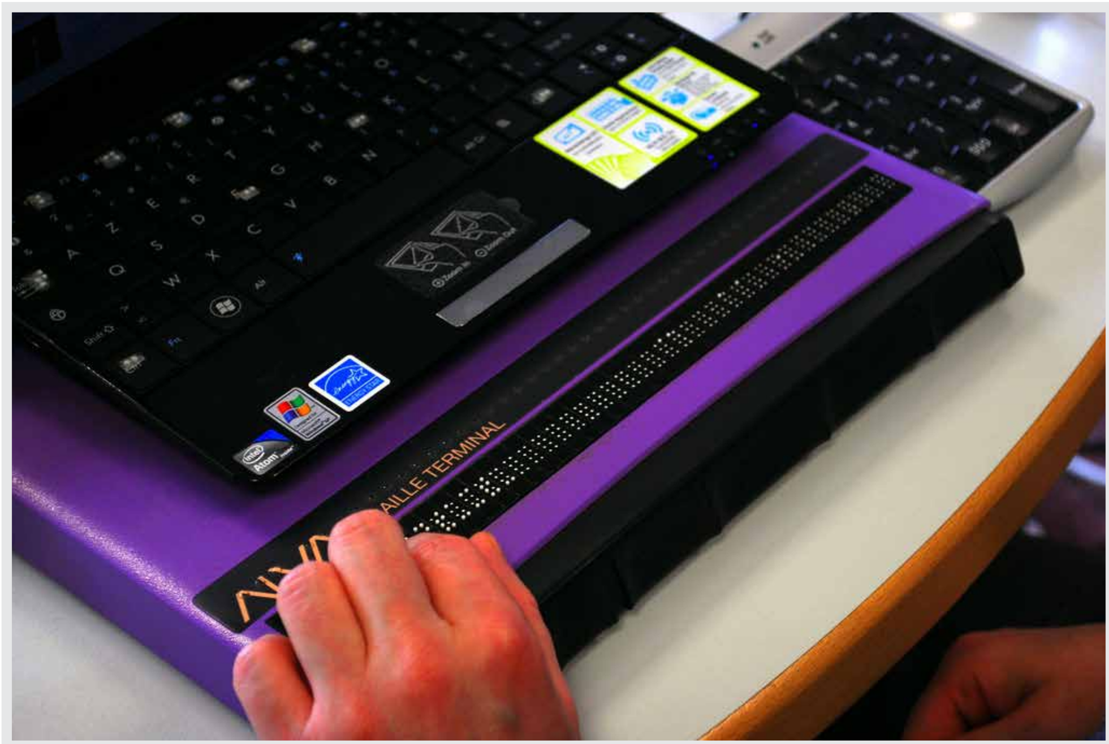




Figure 20.12 A-B: Braille displays.

For example, to achieve the sensation of tracing an unbroken line, a resolution of 20 dpi is required (Fricke and Bahring 1992, Fricke 1997). Even at this resolution, a display area of 20 inches by 15 inches will require 120,000 elements, and yet diagonal lines will still feel broken. Assuming that such technology was available to achieve a higher and more accurate resolution, the technology must be even smaller still. Each of these elements would need to be controlled separately; the system would have to be capable of addressing 120,000 elements individually, but also incredibly quickly. Mechanical technology for this purpose and of this size is simply not available yet, and even the most basic technology that is available is very expensive. One alternative solution to this compound problem has been explored within the Heidelberg Tactile Vision Substitution System

(Maucher et al 2000) which greatly reduces the number of required taxels by using a virtual display area (think of taxels as tactile counterparts to visual pixels). To achieve this, 48 such taxels have been mounted on a carriage that is moved across the larger display area that would have required 2600 taxels; still a very low resolution in terms of overall definition.

### 20.4.2 Haptic Display Technology

Partly influenced by the rapid expansion of the computer games market and general interest in virtual environments, affordable force-feedback and haptic-technology has been available for some time now (force-feedback joysticks, 'rumble-pads', vibro-tactile mice etc.). Based on similar technology to that developed within the Phantom series of force-feedback devices from SensAble Technologies, sophisticated games devices such as the Novint Falcon are being used to offer a rich and immersive sense of physical presence within computer games. Offering three degrees of freedom, a motor controlled arm is used to convey sensations like recoil, impact or different levels of resistance to a player's hand held control device that might take the form of, for example, a gun or bat. The Phantom devices are similar, and like the Falcon can take control of the user's hand or limb, creating a rich variety of virtual textures and other effects. Remember though, haptics implies a combination of both tactile and kinesthetic feedback; however, the devices described here do not really provide much feedback at a tactile level. In this sense, haptic perception within the domain of virtual displays is unlikely to be active and therefore does not match the definition provided earlier. To that extent, the usefulness in such displays lies in their dynamic nature rather than in the richness of haptic interaction that can be achieved.

FIGURE 20.13: A Pair of black Novint Falcons sitting on a table, with the Pistol and Ball grip attachments.



Building on their initial ventures with force-feedback and vibro-tactile feedback gaming devices, the company Immersion is now producing multi-actuator controllers aimed at bringing haptic interaction to the now ubiquitous touch-screen interface on smart-phones and 'tablets'. Designed to be placed behind or around the screen-area, up to 16 actuators can be programmed to produce various vibratory effects that can reinforce user actions. Pressing an onscreen button, for example, can be accompanied by an effect that will go some way towards informing the user that the action has been achieved. Again, this is useful feedback, and the dynamic nature of the display brings considerable flexibility to the interface, but the user, for example, will not be able to feel outline or detailed texture in the way that a static display can offer; the 'display' as such is still predominantly visual.


Figure 20.14: Video gaming devices can make effective use of force feedback to bring immersive qualities to game-play. These range from vibratory ‘rumble’ pads, to simulate collisions and shockwaves for example, to more physical interactions from a force-feedback joystick like the one pictured here (the Sidewinder Force-feedback Pro from Microsoft) where the hand is controlled and moved, perhaps to suggest resistance from obstacles, or actions like the recoil from shooting a gun.

Figure 20.15: More sophisticated force-feedback systems (like those manufactured by Haption) can be used to bring rich haptic-reinforcement into virtual environments.

Figure 20.16: Virtual Reality (VR) parachute trainer. Students wear the VR glasses while suspended in a parachute harness, and then learn to control their movements through a series of computer-simulated scenarios. The computer receives signals from the student as they pull on the risers that control the parachute. The VR trainer also teaches aircrew personnel how to handle a parachute in different weather conditions and during possible equipment malfunctions.



## 20.5 FUTURE DIRECTIONS

It is clear that current technologies cannot yet facilitate dynamic tactile displays that offer the same richness of detail and contrast that we experience with the objects around us on a daily basis. However, as much as this would be an obvious ultimate aim within a context of augmented reality or virtual environments, there are perhaps much more attainable targets that will still be of significant use.

Access to information for visually impaired (VI) users has been a strong feature within our discussion, and it is clear that a truly dynamic and rich tactile display could bring considerable benefits to this area. For example, e-books are now commonplace, but how about e-books where the user can explore and interact with braille and tactile diagrams? Might there be other more mainstream uses for similar technology? At a personal level, my background is shared between technology and music, and though I have considerable enthusiasm for virtual synthesisers and music environments, I still much prefer the sensation of touch and 'feel' with real controls to those afforded by a touch-screen. In this context, which would be more useful to me, a rich colourful and graphical interface with no meaningful tactile interaction or a simple graphical display that allows me to touch and feel the controls? After all, both will be controlling and producing the same sonic output, and that will always be my main concern. Realistically, I imagine that I would probably benefit from a little bit of both. So, how close are we to having such an immersive level of dynamic tactile display? We considered earlier how the size of the mechanical parts creates restrictions in both control and resolution, but there have been significant breakthroughs in this area with the exploration of gel-based pixels that respond to heat. Measuring just 300 microns across, such a pixel will respond to heat from a fine light source and shrink to just half its original height of 0.5mm; the pixel also becomes opaque such that the change is quite visible. The experimental display contains over 4000 of these pixels with about 297 being available within each square centimetre such that fine resolution tactile graphics can be achieved. However, the refresh rate is still slow compared to a standard LCD display, for example, but so is that of a typical e-book. This is still an emerging technology, but one that if developed fully could easily lead to the creation of the tactile e-book that we just considered. If this same technology can be further enhanced to enable input on the part of the user, then this really could be the gateway to enabling rich haptic interaction within all kinds of novel tactile-interfaces.

## 20.6 WHERE TO LEARN MORE

It may come as little surprise that much of the information referenced within this chapter has been drawn from research into non-visual approaches for enabling access to information for visually impaired users. Though there is significant interest in the application of tactile interaction within mainstream devices there is already a considerable body of knowledge that has been established because of research into accessibility through new technologies. There are a number of conferences that address non-visual interaction within a broader range of disability related topics and these include:

- ICCHP - International Conference on Computers Helping People with Special Needs
- CSUN - Annual International Technology and Persons with Disabilities Conference
- ArtAbilitation
- ASSETS - International ACM SIGACCESS Conference on Computers and Accessibility
- ICDVRAT - International Conference on Disability, Virtual Reality and Associated Technologies

Conferences on Human Computer Interaction should also be included as potential sources of further information:

- CHI — ACM Conference on Human Factors in Computing Systems
- HCI — British Computer Society HCI Conference

There are also a number of journals that will occasionally address the notion of Tactile Interaction though perhaps within a wider of interactivity. These include:

- Journal of Visual Impairment and Blindness
- Perception

- Digital Creativity
- Technology and Disability
- International Journal of Arts and Technology
- Computer Music Journal

## 20.7 COMMENTARY BY ALISTAIR D. N. EDWARDS

How to cite this commentary in your report

### Alistair D. N. Edwards

*© Alistair D. N. Edwards*

Alistair Edwards is a Senior Lecturer in Computer Science at the University of York. His research interests are in Human-Computer Interaction (HCI), specifically the use of novel modes in interaction (speech and non-speech sounds, haptics) and interfaces for people with particular needs (e.g. those with visual impairments, older people). Author or co-author of over 70 refereed papers,...

Alistair D. N. Edwards
Alistair D. N. Edwards is a member of The Interaction Design Foundation

Tactile interaction occupies an odd position within the discipline of interaction – as is reflected in this chapter. On the one hand it is associated with futuristic, exotic technology that is still far from mainstream use, but on the other hand it is an element of most conventional human-computer interaction almost without our realizing it. That is to say, that the ubiquitous keyboard and mouse input relies heavily on haptic and proprioceptive senses in an unconscious way that we tend to take for granted. Thus, most people probably do not think that they engage in haptic interaction, but would acknowledge that it must be useful for those who lack other senses (notably sight) for whom tactile communication such as braille would seem invaluable.

It is to be hoped that this chapter will make readers much more aware of the truth of the situation and of the potential of tactile interaction. The keyboard and the mouse have both evolved over time, adapting to the physiological abilities of the user, but with minimal reference to the underlying biology. Of course, the keyboard is a classic example of non-optimal evolution: it is suggested (Noyes 1983) that the conventional *qwerty* layout was designed to slow typists down, to reduce consecutive-letter clashes on mechanical typewriters and now we are stuck with it. On the one hand, it is fun to speculate how the keyboard might have been designed if the inventors had had the benefit of reading a chapter such as this, but on the other, one has to note that even with the availability of this knowledge haptic interaction clearly a long way from reaching its full potential.

We have physical devices, including keyboards and mixing desks and lighting desks, which inherently give good haptic feedback and interaction, but in the digital, virtual world, we are constrained by the technology. Is the investment (of time, money, space etc.) required to use a device such as that depicted in Figure 21.13 worth the benefit derived?

The tactile picture is a device which seems in some sense 'obvious', particularly to those of us who have sight: if you can't see a picture, then you should be able to feel it. I have spent a lot of time and effort working with tactile diagrams, mainly within the University of York Centre for Tactile Images (sadly no longer in existence for lack of financial viability). We produced a wide range of diagrams for different purposes, working in collaboration with a number of important clients (Including *The Deep*, Hull; The Jorvik Centre; The National Trust; English Heritage and The National Railway Museum). Repeatedly we found ourselves repeating the assertion (mentioned by Challis) that a good tactile diagram may *look* bad. That is to say that the haptic senses are very different from vision.

The immediate differences are obvious – when one stops to think about them. Tactual sensitivity is low. Challis mentions the two-point limen but refrains from giving a figure for it. This is probably wise, since it is at least controversial as a measure of tactual sensitivity and is hard to gather agreement as to average values. The point is, though, that the value is of the order of millimetres – much lower than visual resolution. In Figure 21.5, Challis shows examples of patterns which could be produced on swell paper. In the visual representation they are all clearly distinct, but in experiments with swell paper samples (Magill 1999), we found that the average person can only distinguish three levels of pattern: smooth, medium and rough, regardless of their visual pattern. For instance, while the vertical and horizontal cross-hatching shown in Figure 21.5 are clearly different visually, they would probably be perceived as being the same on tactual inspection.

Furthermore, in exploring a tactile diagram, the person is likely to use at most two finger tips and more likely one. It is easy to characterize this as exploring through a tactile pinhole. Yet, in practice, it is likely that the situation is even worse than that. Vision is an inherently integrative sense. The angle sensed by the eye is very small, but because the eye moves constantly and the brain can integrate information we get a wide field of view which forms a picture that we perceive as coherent. The only people who need to be taught how to perceive visually are those who have lacked the sense of vision and have then had it given to them (usually through surgery, as in the case of Virgil in (Sacks 2009)), whereas even proficient braille readers need to be taught how to use tactile pictures. Evidence from such cases of lately-acquired sight suggests that it may be that people who mature without sight may never develop the ability to integrate fragments into a comprehensive picture (literally and metaphorically). In other words, a tactile picture may always be of limited value – even if it has been designed in such a way as to make the most of the haptic senses.

A limitation of most tactile graphic technologies, such as swell paper, is that it is static and hence the idea of a tactile screen is attractive – and perennial. I have stopped paying much attention to announcements and papers by people who have had this idea again and believe they have found the one which will work. I would like to propose a moratorium on such publications until someone can present their device which works, is reliable and affordable. In the absence of such a ban, I will impose my own by not reading them. I have no doubt that one day someone will invent the appropriate technology, but – for the reasons set out in this chapter – it is a hard goal to achieve. Even this chapter mentions another of these attempts ('gel-based pixels that respond to heat'), but clearly it is quite speculative again, since there is not even a publication cited.

It is clear that the haptic sense becomes even more important in the absence of the dominating sense of vision. Braille is the richest form of tactile communication its use by blind people is very low. It is estimated that fewer than 2% of blind people read braille (Bruce, McKennell et al. 1991). The reasons for the low take-up are unclear and probably complex, but the essential explanation must be that blind people do not perceive that the benefits they will gain are worth the effort that it will take to learn braille. This particularly applies to those who have had sight and have lost it – which is the vast majority. For those of us who have sight, exclusion from (printed) literature seems a great loss, but for most of those who experience this the difficulty of learning to read tactually is not sufficient motivation.

Yet it is not sheer laziness on the part of sighted people that means that they do not make as much use of their non-visual senses as they might. In radical experiments by (Pascual-Leone, Theoret et al. 2004) sighted participants were blindfolded 24 hours per day for several days and given daily lessons in reading braille. They made good progress in learning braille and brain scans showed that areas of the brain cortex normally associated with visual processing were being reassigned to processing tactile information. A control group of non-blindfolded, sighted participants made minimal progress on learning braille and showed no brain adaptation. In other words it seems that braille is not only likely to be useful for blind people, they are the only people who are likely to be able to use it.

To summarize, haptic senses are used to a greater extent in conventional interaction than most people realize. They could be used more and to better effect if designers were more informed regarding them. Virtual haptics are being used increasingly and this trend will accelerate as the technology develops, but there is something to be said for the use of (well-designed), old-fashioned physical interactors (knobs, switches and the like). Ultimately, though, there is a limit to the use that can be made, given the physiological limitations of the haptic senses.

### 20.7.1 References


- Bruce, I., A. McKennell, et al. (1991). Blind and Partially Sighted Adults in Britain: The RNIB Survey. London, HMSO.
- Magill, D. (1999). Experiments to test differentiation of patterns for tactile human-computer interaction MSc (IT) Project Report, Department of Computer Science, University of York.
- Noyes, J. (1983). "The QWERTY keyboard: a review." International Journal of Man-Machine Studies **18**(3): 265-281 DOI: http://dx.doi.org/10.1016/S0020-7373(83)80010-8.
- Pascual-Leone, A., H. Theoret, et al. (2004). Tactile processing in the visual cortex. Touch Blindness and Neuroscience. S. Ballesteros and M. A. Heller. Madrid, Spain, UNED: 51-52.
- Sacks, O. (2009). An Anthropologist on Mars, Picador.

## 20.8 COMMENTARY BY KLAUS MIESENBERGER

How to cite this commentary in your report

### Klaus Miesenberger

© *Klaus Miesenberger*

Klaus Miesenberger is vice head of Institute Integriert Studieren at the University of Linz, Austria. He has a background computer science and economics. He is responsible for R&D and teaching at the institute which also runs a service centre for students with disabilities. In 2000 he was guest professor at the Université Claude Bernard, Lyon II. He gives lectures at different Austrian u...

Klaus Miesenberger
Klaus Miesenberger is a member of The Interaction Design Foundation

I like the chapter a lot — a brilliant source for students and practitioners in need of the state of the art. Two things in particular:

1. We must not forget to look at the fundamentals, results, success stories and failures before jumping into a new adventure!

2. Things which are difficult to grasp makes it hard for readers and students to cope with the subject. This chapter helps by summarizing and bringing the issue to the point.

The chapter presents the different components, the bits and pieces relevant for tactile information presentation, the potential of tactile/tangible/touch interfaces and the challenges we are facing on the way to better support efficiency, quality and joy of interaction with technical systems and content. The chapter does not promise what it can't answer and does not try to answer what the next steps in tactile/tangible/touch interaction will be — as nobody could foresee that touch displays have suddenly come to proliferate: Technology, concepts, ideas and prototypes were available, described and discussed for decades, but only recently the concept and its implementations exploded. Why has tangible interaction been so successful in attracting a big part of the "non-techy" population and get so many people — including aging people and people with disabilities — engaged with ICT? Is it touch?, is it vision?, is it overcoming of hand/touch-eye-ear- coordination?, is it the mobile aspect?, is it non-text oriented aspect?, ...? In the end, many answers and factors have to be taken into account to grasp the big picture allowing us a next step in the journey of interface design e.g. by adding a new quality of perception into the tactile experience.

Ben Challis' chapter collects what might form a next step in the interface revolution. This revolution could be summarized as "From the hand-axe to the pointer": It started with switches and switchboards (with lots of tangible interaction), over console, WIMP, SILK towards tactile/touch/tangible. Interesting to see that first we moved away from immediate and intuitive interaction using tactile interaction towards a new virtual, intangible experience (i.e. the WIMP interface) with its advantages in terms of efficiency and independence, but also with its limitations in acceptance due to a lack of intuitiveness.

In our pursuit of creating user experiences that are intuitive, usable, enjoyable and efficient, research and development increasingly uses touch and sound as well as other senses like acceleration, gravity and the olfactory sense to achieve this. Tactile interaction combines several types of sensory experiences, which is perhaps the reason it seems intuitive — it give us a feeling of easy, efficient, effective, straight forward and joyful interaction. As designers, we must however first understand touch, tactile and tangible interfaces on their own before advancing into cross-sensing approaches.

Ben's chapter provokes such questions, which makes it an ideal source for students, researchers and developers when diving into what seems "obvious" at first glance. We must never expect simple, easy to apply answers in interface and interaction design. We have to learn to put the right questions in the right context, to be prepared to question the obvious and to progress based on a sound and respectful analysis of the state of the art. This is what this chapter is.

And last but not least, I like how we learn from the extreme: People with reduced or no visual sense in many aspects show how tactile or alternative interfaces have to be designed. They are key experience and know-how owners.

Good interfaces "don't make us think". But good interface design depends on designers and developers willing and prepared to think about the big picture of interaction and interface design. Ben's chapter is the right source to make us think before we implement!

## 20.9 REFERENCES

*Bentzen, Billie Louise and Peck, Alec F. (1979): Factors Affecting the Traceability of Lines for Tactile Graphics. In Journal of Visual Impairment and Blindness, 73 (7) pp. 264-269*

*Berla, Edward P. (1972): Behavioral Strategies and Problems in Scanning and Interpreting Tactual Displays. In New Outlook for the Blind, 66 (6) pp. 277-286*

*Berla, Edward P. and Butterfield, Lawrence H. (1977): Tactual Distinctive Features Analysis: Training Blind Students in Shape Recognition and in Locating Shapes On a Map. In The Journal of Special Education, 11 (3) pp. 335-346*

*Berla, Edward P. and Murr, Marvin J. (1975): Psychophysical functions for active tactual discrimination of line width by blind children. In Perception and Psychophysics, 17 (6) pp. 607-612*

*Challis, Ben (2002): Designing Interactive Tactile Diagrams. In: Miesenberger, Klaus, Klaus, Joachim and Zagler, Wolfgang L. (eds.) ICCHP 2002 - Computers Helping People with Special Needs - 8th International Conference July 15-20, 2002, Linz, Austria. pp. 559-561*

*Challis, Ben and Edwards, Alistair D. N. (2000): Design Principles for Tactile Interaction. In: Proceedings of the First International Workshop on Haptic Human-Computer Interaction 2000. pp. 17-24*

*Challis, Ben and Edwards, Alistair D. N. (2000): Weasel: A System for the Non-Visual Presentation of Music Notation. In: Proceedings of 6th International Conference on Computers Helping People with Special Needs. ICCHP 2000 2000.*

*Fricke, Joerg and Baehring, Helmut (1993): Design of a tactile graphic I/O tablet and its integration into a personal computer system for blind users. In Journal of Microcomputer Applications, 16 (3) pp. 259-269*

*Fricke, Joerg and Bahring, H. A. (1992): A Graphic Input/Output Tablet for Blind Users. In: Proceedings of 3rd International Conference on Computers for Handicapped People 1992.*

*Fricke, J. and Edwards, Alistair D. N. (2000): Tactile Displays Based on Modulated Electromagnetic Radiation. In: Proceedings of 6th International Conference on Computers Helping People with Special Needs. ICCHP 2000 2000.*

*Heller, Morton A. (1992): Haptic dominance in form perception: vision versus proprioception. In Perception, 21 (5) pp. 655-660*

*James, G. A. and Gill, J. M. (1975): A pilot study on the discriminability of tactile areal and line symbols for the blind. In Research Bulletin of the American Foundation for the Blind, 29 pp. 23-31*

*Klatzky, Roberta L. and Lederman, Susan J. (1987): The Representation of Objects in Memory: Contrasting Perspectives from Vision and Touch. In: Gruneberg, M. M., Morris, P. E. and Sykes, R. N. (eds.). "Practical Aspects of Memory: Current Research and Issues". Wiley*

*Lambert, L. M. and Lederman, Susan J. (1989): An evaluation of the legibility and meaningfulness of potential map symbols. In Journal of Visual Impairment and Blindness, 83 pp. 397-403*

*Lederman, Susan J. (1982): The Perception of Texture by Touch. In: Schiff, William and Foulke, Emerson (eds.). "Tactual Perception: A Sourcebook". Cambridge University Press*

*Lederman, Susan J. and Campbell, J. I. (1983): Tangible line graphs: An evaluation and some systematic strategies for exploration. In Journal of Visual Impairment and Blindness, 77*

*Lederman, Susan J. and Kinch, Denise H. (1979): Texture in Tactual Maps and Graphics for the Visually Handicapped. In Journal of Visual Impairment and Blindness, 73 (6) pp. 217-227*

*Loomis, Jack M. (1981): Tactile pattern perception. In Perception, 10 (1) pp. 5-27*

*Loomis, Jack. M. and Lederman, Susan J. (1986): Tactual Perception. In: Boff, Kenneth R. (ed.). "Handbook of Perception and Human Performance". New York: John Wiley and Sons*

*Maucher, Thorsten, Schemmel, Johannes and Meier, Karlheinz (2000): The Heidelberg Tactile Vision Substitution System. In: Proceedings of 6th International Conference on Computers Helping People with Special Needs, ICCHP 2000 2000.*

*McDonnell, Paul M. and Duffett, James (1972): Vision and touch: a reconsideration of conflict between the two senses. In Canadian Journal Of Psychology, 26 (2) pp. 171-180*

*Millar, Susanna (1985): The perception of complex patterns by touch. In Perception, 14 (3) pp. 293-303*

*Nolan, Carson Y. (1971): Relative Legibility of Raised and Incised Tactual Figures. In Education of the Visually Handicapped, 3 (2) pp. 33-36*

*Nolan, Carson Y. and Morris, June E. (1971): Improvement of tactual symbols for blind children: 1 June 1964--28 February 1969 : final report. American Printing House for the Blind*

*Oakley, Ian, McGee, Marilyn Rose, Brewster, Stephen A. and Gray, Philip D. (2000): Putting the Feel in 'Look and Feel'. In: Turner, Thea, Szwillus, Gerd, Czerwinski, Mary, Peterno, Fabio and Pemberton, Steven (eds.) Proceedings of the ACM CHI 2000 Human Factors in Computing Systems Conference April 1-6, 2000, The Hague, The Netherlands. pp. 415-422*

*Richter, Andreas and Paschew, Georgi (2009): Optoelectrothermic Control of Highly Integrated Polymer-Based MEMS Applied in an Artificial Skin. In Advanced Materials, 21 (9) pp. 979-983*

*Rock, Irvin and Victor, Jack (1964): Vision and Touch: An Experimentally Created Conflict between the Two Senses. In Science, 143 (3606) pp. 594-596*

## About the author

### Ben Challis

© *Ben Challis*

Has also published under the name of:
"Ben P. Challis"

Ben Challis is a composer, performer and technologist. With research interests that embrace the notion of design-for-all within music-performance, he has worked on various projects that explore alternative modes of interaction with sound and music for people with specific individual needs. As a performer, he works with these same technologies, exploring their creative and expressive potential within free-improvisation. As composer he has composed scores for film and theatre productions. He is a Senior Lecturer and Joint Award Leader in Popular Music at the University of Glamorgan, ATRiuM.

## Your notes and thoughts on chapter 20

Record your notes and thoughts on this chapter. If you want to share these thoughts with others online, go to the bottom of the page at:
http://www.interaction-design.org/encyclopedia/tactile_interaction.html

**NOTES**

# Chapter 21

# Somaesthetics

## Thinking Through the Body and Designing for Interactive Experience

*by Richard Shusterman.*

What is somaesthetics and why should it appear as a core article in an encyclopedia devoted to Human-Computer Interaction and Interaction Design? Readers trained in design studies and informational technology should not feel guilty if the term strikes them as unfamiliar and even difficult to decipher or pronounce. As a new interdisciplinary field whose roots are in philosophical theory, somaesthetics offers an integrative conceptual framework and a menu of methodologies not only for better understanding our somatic experience, but also for improving the quality of our bodily perception, performance, and presentation. Such heightened somatic awareness and mastery offers benefits to many fields including design. Our experience of ourselves and our world is always embodied and involves somatic responses and feelings that are typically unnoticed though they are unavoidable and indispensable for our proficient functioning. We need a proper feel for our tools in order to use them most effectively; and this includes the use of one's own body in using other tools. For the body is our indispensable tool of tools, the necessary medium of our being, perception, action and self-presentation in the world. By exploring the fundamental features of our embodied ways of engaging the world and transforming it through action and construction, somaesthetics can provide useful insights and experiential skills to help designers produce products and situations that provide more rewarding and pleasurable experience.

Though somaesthetics is grounded in philosophical theory and therefore does not command the wide media attention and advertising hype that technological inventions often receive, it is not a narrowly abstract discipline that advocates pure theory over practice and concrete applications. Nor does it hide behind abstruse technical jargon. It emerges from American pragmatist philosophy that insists on the primacy of practice even in the constructing and testing of theories and that equally insists on clear language, empirical evidence, and practical results. Somaesthetics reflects the pragmatist idea of philosophy as a means of improving experience through a reflective art of living. Philosophy, in this sense, is a tool for designing life.

Somaesthetics is an interdisciplinary research product devoted to the critical study and meliorative cultivation of the experience and use of the living body (or soma) as a site of sensory appreciation (aesthesis) and creative self-stylization. An ameliorative discipline of both theory and practice, somaesthetics seeks to enrich not only our discursive knowledge of the body but also our lived somatic experience and performance; it aims to improve the meaning, understanding, efficacy, and beauty of our movements and of the environments to which our actions contribute and from which they also derive their energies and significance. To pursue these aims, somaesthetics is concerned with a wide diversity of knowledge forms, discourses, social practices and institutions, cultural traditions and values, and bodily disciplines that structure (or could improve) such somatic understanding and cultivation, and it is therefore an interdisciplinary project, in which theory and practice are closely connected and reciprocally nourish each other. It is not limited to one theoretical field, academic or professional vocabulary, cultural ideology, or particular set of bodily disciplines. Rather it aims to provide an overarching theoretical structure and a set of basic and versatile conceptual tools to enable a more fruitful interaction and integration of the very diverse forms of somatic knowledge currently being practiced and pursued. There is an impressive, even overwhelming abundance of discourse about the body in many disciplines of contemporary theory and commercial enterprise. But such somatic discourse typically lacks two important features. First, a structuring overview or architectonic that could integrate their very different discourses into a more productively coherent or interrelated field. It would be useful to have a broad framework (which does not mean a unified, highly consistent system) that could connect, for example, the discourse of biopolitics to the therapies of bioenergetics, the neuroscience of hand gestures to their aesthetic meaning in Nõ theater. The second feature lacking in most academic discourse on embodiment is a clear pragmatic orientation — something that the individual can clearly employ or apply to his or her life in terms of disciplines of improved somatic practice. Somaesthetics offers a way to address both these deficiencies.

## 21.1 GENEALOGY AND EMERGENCE

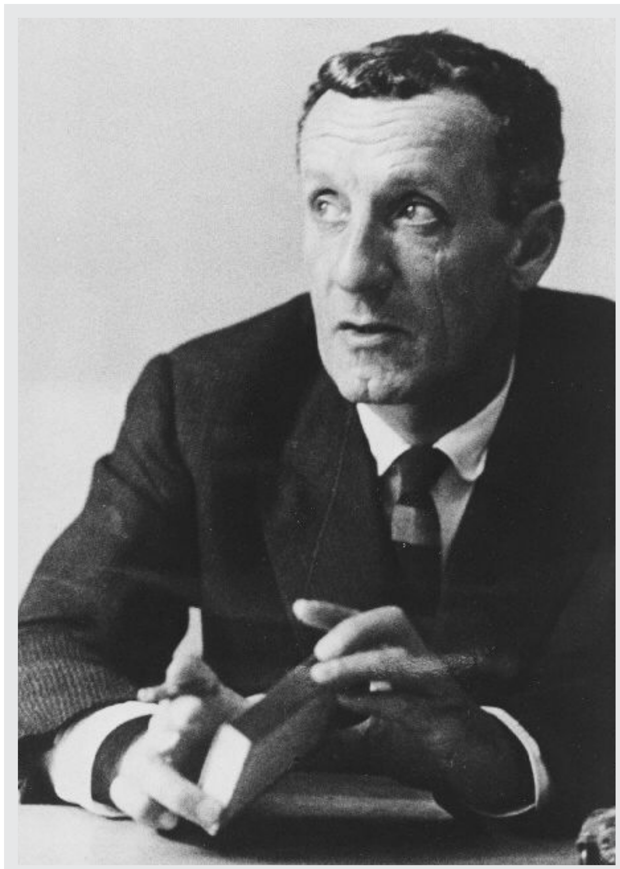


FIGURE 21.1: The French philosopher Maurice Merleau-Ponty (1908-1961) who affirmed the body as the centre of human cognition.



The research project of somaesthetics delineated above began to emerge in the mid-1990s from two principal themes in my research: pragmatist aesthetics and philosophy as an embodied art of living. If pragmatist aesthetics rejects the traditional aesthetic attitude of distanced, disinterested contemplation by advocating an aesthetics of active, creative engagement, then it also should recognize that all action (artistic, practical, or political) requires the body, our tool of tools. Building on the pragmatist insistence on the body's central role in artistic creation and appreciation, somaesthetics highlights and explores the soma — the living, sentient, purposive body — as the indispensable medium for all perception. If experiences of art and beauty are distinctive for the powerfully gratifying ways they absorb our attention, unify our consciousness, and engage our emotions, then increasing our powers of awareness, focus, and feeling through better mastery of their somatic source could render more of our experience similarly rewarding in such ways. Not only art's creation and appreciation would be enhanced through this heightening of consciousness, but also the attractive shaping of our lives as an art of living could be enriched by greater perceptual awareness of aesthetic meanings, feelings, and potentials in our everyday conduct of life.

FIGURE 21.2: Confucius (551-479 BC), the Chinese philosopher, who advocated cultivation and refinement of the body, particularly through art and ritual.

This ancient pragmatic idea of philosophy as a way of life or art of living is thus a second root of the somaesthetic project, with its integration of theory and practice. Embodiment has become an increasingly trendy theme in academia but somaesthetics takes the notion of embodied philosophy in a distinctively strong sense, even stronger than that of phenomenologies like Merleau-Ponty's in which the body forms a central perspective that structures the philosophical system and is celebrated as a sentient, intelligent, purposive, skilled subjectivity that likewise helps construct the world rather than being a mere physical object in it. Rather than the phenomenological quest to reveal an alleged primordial, foundational, and universal embodied consciousness that (in Merleau-Ponty's words) is "unchanging, given once and for all," and "known by all men" in all cultures and times, somaesthetics (as I conceive it) recognizes that somatic consciousness is always shaped by culture and thus admits of different forms with different cultures (or with different subject positions within the same culture)[1]. Second (and more important than , somaesthetics is interested not merely in describing our culturally shaped forms of somatic consciousness and modes of somatic practice but also in improving them. Third, to effect such improvements, it also includes practical exercises of somatic training rather than mere philosophical discourse[2].

Building on pragmatic insights and ancient philosophical traditions from both the East and the West, somaesthetics advocates somatic training as a wor-

1. See Maurice Merleau-Ponty, *Phenomenology of Perception*, trans. Colin Smith(London: Routledge, 1962), xiv; and Maurice Merleau-Ponty, *In Praise of Philosophy and Other Essays* , trans. John Wild, James Edie, and John O'Neill (Evanston:Northwestern University Press, 1970), 63. Of course, somaesthetics is open enough not to preclude the possibility of fundamental, universal forms of embodied being in the world and thus is tolerant enough to admit such inquiries within its scope. My personal view, however, is that there is little hope of extracting such a substantive primordial, universal somatic consciousness that is the same for all times, all ages, all cultures, and all genders. Our somatic experience, it seems to me, is always already thoroughly shaped by culture as well as nature. In that sense, human nature is intrinsically cultural. For my arguments on these points, see Richard Shusterman, Body Consciousness: A Philosophy of Mindfulness and Somaesthetics (Cambridge: Cambridge University Press, 2008).

2. What precisely constitutes improvement is not a question that admits of a single, general, definitive answer. Different contexts and problems will demand different solutions. Moreover, one dimension of somaesthetic inquiry involves the debate over somatic norms, methods, and values that eventually determine how to understand improvement in particular contexts.

thy dimension of philosophical cultivation and expression. Confucius clearly affirmed somatic cultivation as a crucial dimension of philosophical education, once informing his disciples that he could cease speaking and simply teach as nature does by embodying his philosophy in his bodily behavior. Daoism also advocated somatic cultivation, though sometimes in very different modalities. For more detailed discussions of Confucian and Daoist views relating to issues in somaesthetics, see Shusterman (2009). Greek and Roman thinkers often likewise advocated this ideal, sometimes by contrasting true philosophers who lived their philosophy to those who merely wrote philosophy and thus were denigrated as mere "grammarians[3]." Invoking this ancient tradition, I thus described somaesthetics "as a new name for some old ways of thinking," borrowing the shrewd formulation William James used to subtitle his first book on pragmatism. I used the term "soma" (a less familiar expression deriving from the Greek word for body) to avoid problematic associations of body (which can be a lifeless, mindless thing) and flesh (which designates only the fleshly parts of the body and is strongly associated with Christian notions of sin) and to insist that my project concerns the lived, sentient, purposive body rather than merely a physical body. The "aesthetic" in somaesthetics also originates in Greek, deriving from the word for sensory perception (*aisthesis*) that Alexander Baumgarten used to coin the modern philosophical discipline of aesthetics in 1758. So "somaesthetics" (a simple splicing

3. Philosophy, of course, in its fullest sense, surely should include practices of theoretical writing as well as living. As I write in *Practicing Philosophy*, "Though one may usefully distinguish between philosophy as theory and as artful living — between books and life one must not erect this into a false dichotomy. First, writing is...an important tool for artfully working on oneself — both as a medium of self-knowledge and of self-transformation...Secondly, philosophical theories of the world typically serve as logical grounds or guiding orientations through which philosophical arts of living are developed and defended. ...The point I am making is that there is no essential opposition compelling us to choose between philosophy as theory and as artful life-practice. Indeed, we must *not* choose between them. For...we surely should build our art of living on our knowledge and vision of the world, and reciprocally seek the knowledge that serves our art of living." Somaesthetics is devoted to such integration of theory and practice in our approach to embodiment. For more on these points, see Richard Shusterman, *Practicing Philosophy: Pragmatism and the Philosophical Life* (London: Routledge, 1997); citations from 3-4; and "Pragmatism and East-Asian Thought," in Richard Shusterman (ed.), *The Range of Pragmatism and the Limits of Philosophy* (Oxford: Blackwell, 2004).

of “soma” and “aesthetics”) implies a project of appreciating and cultivating the body not only as an object that externally displays beauty, sublimity, grace, and other aesthetic qualities, but also as a subjectivity that perceives these qualities and that experiences attendant aesthetic pleasures somatically. When Baumgarten founded aesthetics, he intended it as a general science of sensory cognition rather than a field exclusively devoted to art and objects of beauty. But though his theory was focused on the senses, he excluded all considerations of the bodily dimension of sensation, believing that sensory perception was an entirely mental affair. Somaesthetics thus returns aesthetics to its roots as a science of sensory perception but insists on the somatic dimension of such perception and of our action, thought, and feeling as well.

From the outset, somaesthetics has had an international career. I first introduced it in a 1996 book in German, *Vor der Interpretation*, using the term “Somästhetik” to designate this project, before its initial English presentation in *Practicing Philosophy: Pragmatism and the Philosophical Life* (1996), and then the first detailed articulation of its structure in “Somaesthetics: A Disciplinary Proposal” (1999, reprinted in *Pragmatist Aesthetics*, 2nd edition, 2000). *Performing Live* (2000) represents a further stage in the development of somaesthetics in which its connection with the new media is discussed and different methodologies for heightening body consciousness are analyzed, while *Body Consciousness* (2008) constitutes my most comprehensive treatment of somaesthetics, though it focuses primarily on the experiential dimension of the somaesthetic field, whose general structure is outlined in the section 21.3 of this article.

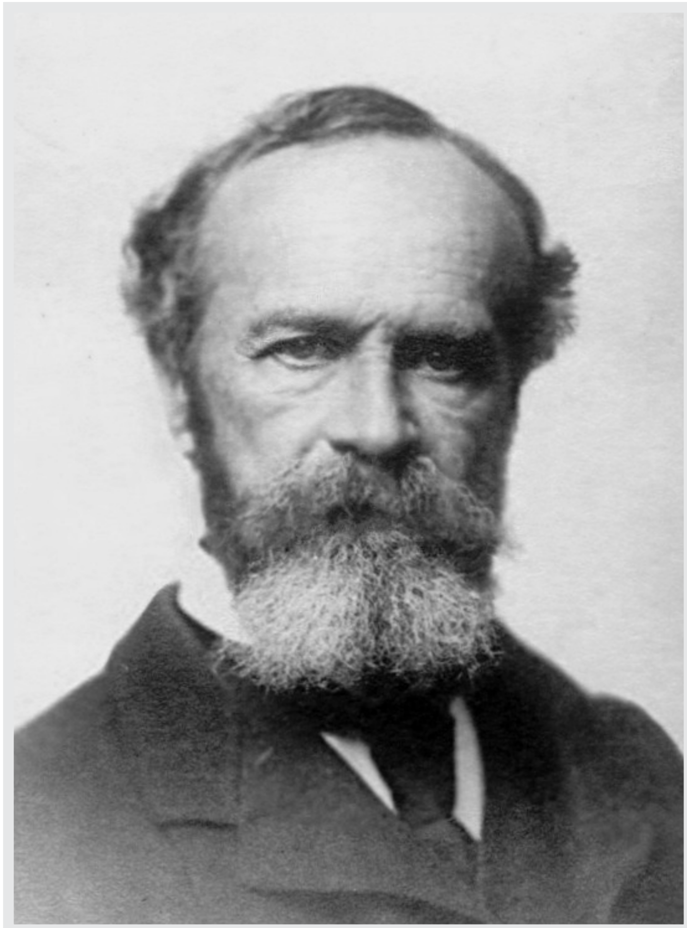
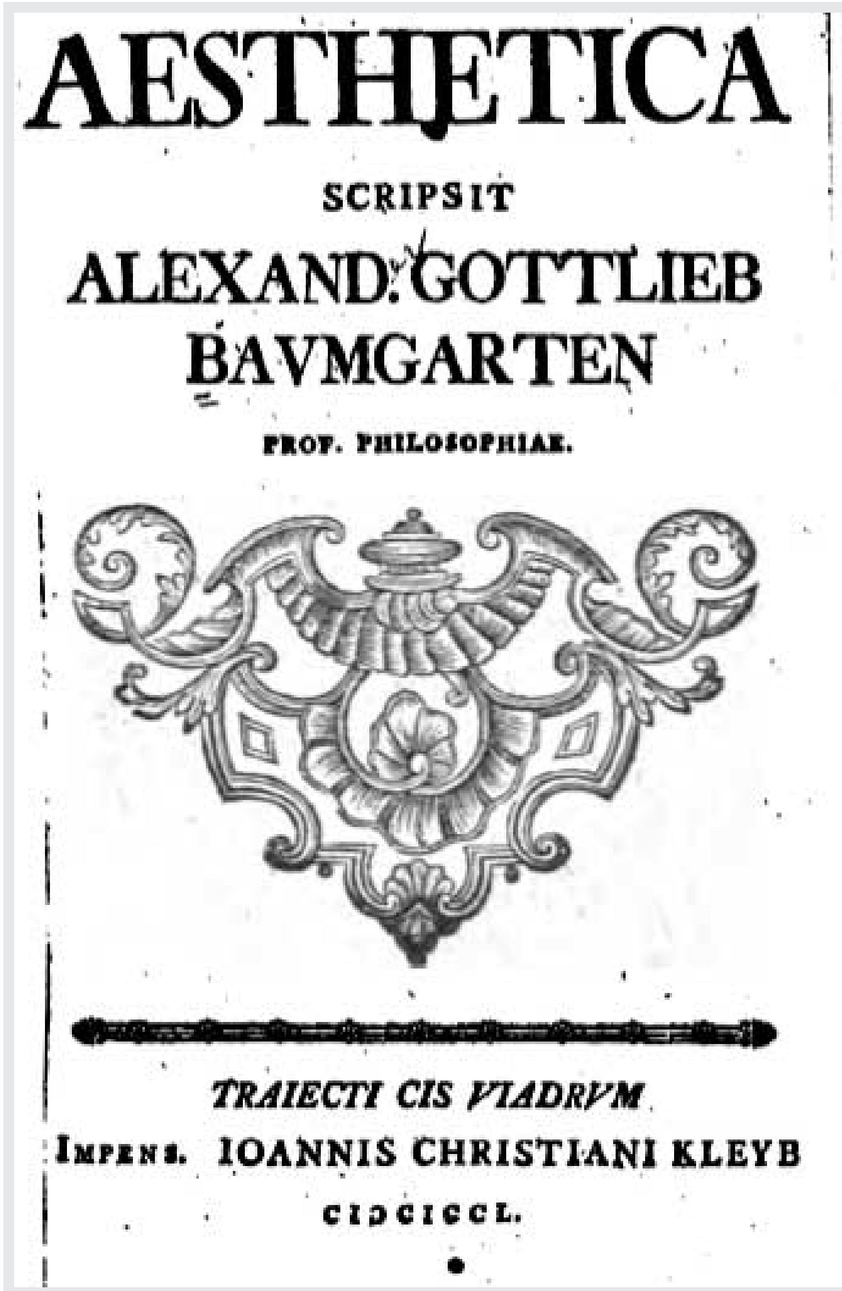


FIGURE 21.3.A: The American psychologist and philosopher William James (1842-1910), whose pragmatist philosophy has inspired somaesthetics and its insistence on the interaction between theory and practice.

*Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.*



AESTHETICA

SCRIPSIT

ALEXAND. GOTTLIEB

BAVMGARTEN

PROF. PHILOSOPHIAE.

TRAIECTI CIS VIADRVM

IMPENS. IOANNIS CHRISTIANI KLEYB

CIↃCICCL.

FIGURE 21.3.B: In his book Aesthetica from 1750, the German philosopher Alexander Gottlieb Baumgarten (1714-1762) coined the philosophical discipline of aesthetics that rehabilitates the cognitive powers of the senses.

## 21.2 SOMAESTHETICS AS AN INTERDISCIPLINARY FIELD

FIGURE 21.4: Somaesthetics as performance art and photography: From the Soma Flux series, photographed by Yann Toma at Cartagena, Colombia.



Initially I thought of the somaesthetic project as being fully nested within the discipline of philosophy, perhaps as a branch of aesthetics. But I soon realized that somaesthetics should be an essentially interdisciplinary field, even if grounded in philosophy. As all human perception and action goes through the soma, many different academic disciplines can contribute significantly to the study and improvement of somatic experience and performance. It would be foolish, therefore, to limit somaesthetics to the methods and concerns of philosophy. Engaging a wide variety of knowledge forms and disciplines that structure our somatic experience or can improve it, somaesthetics is a framework to promote and integrate the diverse range of theorizing, empirical research, and melioriative practical disciplines concerned with bodily perception, performance, and presentation. While originally rooted in my philosophical research, it is not a single theory or method advanced by a particular philosopher but an open field for collaborative, interdisciplinary, and transcultural inquiry. Its applications already extend beyond philosophy to a broad array of topics ranging from the arts, product design, and politics to fashion, health, sports, martial arts, and the use of hallucinogenic drugs in education[4]. Somaesthetics' most notable developments thus far can be grouped into three general areas: arts, politics, and design of technology.

4. See, for instance, TittiKallio, "Why we choose the more attractive looking objects: somatic markers and somaesthetics in user experience," in *Proceedings of the 2003 International Conference on Designing Pleasurable Products and Interfaces*, ACM, 142-143; N.W. Loland, "The Art of Concealment in a Culture of Display: Aerobicizing Women's and Men's Experience and Use of Their Own Bodies, *Sociology of Sport Journal*, 17 (2000), 111-129; J. G. Forry, "Somaesthetics and Philosophical Cultivation: An Intersection of Philosophy and Sport," *ActaUniversitatisPalackianaeOlomucensis. Gymnica*, Vol 36, No 2 (2006), 25-28; Michael Surbaugh, "'Somaesthetics,' Education, and Disabilty," *Philosophy of Education*, 2009, 417-424; S.J. Smith and R.J. Lloyd "Promoting Vitality in Health and Physical Education," *Qualitative Health Research*, 16:2, (2006) 249-267; Ken Tupper, "Entheogens and Education," *Journal of Drug Education and Awareness*, 1:2 (2003), 145-161;

FIGURE 21.5: Somaesthetics as performance art and photography: From Yann Toma's Soma Flux series at Cartagena, Colombia.



Though dance may be the most paradigmatic of somatic arts, somaesthetics has been equally applied to theatre in analyzing the somatic styles of movement and posture of actors on stage[5].

5. For dance, see

- Peter Arnold, "Somaesthetics, Education, and the Art of Dance," *Journal of Aesthetic Education*, 39 (2005), pp. 48—64.
- Bryan Turner, "Introduction — Bodily Performance: On Aura and Reproducibility," *Body and Society*, 11:4 (2005), pp. 1—17;
- Lis Engel, " The somaesthetic dimension of dance art and education - a phenomenological and aesthetic analysis of the problem of creativity in dance," in E Anttila, S Hämäläinen, T Löytönen & L Rouhiainen (eds), *Ethics and politics embodied in dance: Proceedings of*

Somaesthetic concepts and theories have been even more extensively deployed for understanding music and music education[6]. In visual arts, somaesthetics has been used to explain not only how artists use their bodies in making artworks but also how observers deploy themselves somatically to perceive such works. Many works of visual art (whether paintings, sculptures, photographs, or installations) consciously presuppose and play with the viewers' somatic standpoint, so that the soma can be powerfully thematized in a work without a body being visually represented in it[7]. The body (with its multiple senses and movement through space) likewise plays a formative role in architectural design and experience. Performance art presents a distinctive case in which the body is not only a tool of creation and means of perception but also the expressive medium and visual end-product or art object. Building on my somaesthetic theory, Martin Jay shows the political import of body-centered performance works that challenge the prevailing norms of bodily form and comportment with their attendant sociopolitical hierarchies of domination[8].

Somaesthetics has begun to have an impact not only on the analysis of visual art, but also on its practice. One prominent example is its use as a generative theo-

---

*the International Dance Conference*, December 9-12, 2004 (Helsinki: Theatre Academy, 2005), 50-58;

- Patricia Vertinsky "Transatlantic Traffic in Expressive Movement: From Delsarte and Dalcroze to Margaret H'Doubler and Rudolf Laban," *The International Journal of the History of Sport*, 26: 13 (2009), 2031 — 2051.

For theatre, see

- Eric Mullis, "Performative Somaesthetics: Principles and Scope," *Journal of Aesthetic Education*, 40 (2006), 104-117,
- and with respect to Japanese theatre, see my paper "Body Consciousness and Performance: Somaesthetics East and West, *Journal of Aesthetics and Art Criticism*, 67.2 (2009), 133-145.

6. See, for example, the special issue on somaesthetics (as presented by my book *Body Consciousness*) in the journal, *Action, Criticism, and Theory for Music Education*, 9:1 (2010), http://act.maydaygroup.org/php/archives_v9.php#9_1
7. For applications of somaesthetics to the visual arts, see, for example, David Zerbib, "Soma-esthetique du corps absent," in Barbara Formis (ed.), *Penser en corps. Soma-esthetique, art, et philosophie* (Paris: L'Harmattan, 2009), 133-159; Aline Caillet, "Emanciper le corps: sur quelques applications du concept de la soma-esthétique en art," in Formis (ed.), 99-112.
8. See Martin Jay, *Refractions of Violence* (New York: Routledge, 2003), 163-176.

retical background for Peng Feng's curatorial project for the Chinese Pavilion of the 2011 Venice Biennale. Entitled *Pervasion*, this show of five installation pieces (including clouds with tea fragrance; pipes dripping with Chinese schnapps; fragrant porcelain pots of herb medicine; fog of incense; and lotus-scented virtual snow) sought to emphasize that our appreciation of even visual art is always much more than visual and to highlight the soma's role as transmodal perceiving subjectivity by engaging also the pleasures of other bodily senses[9]. Somaesthetics has also been used as a creative framework for a series of photographic and cinematic works that the Parisian artist Yann Toma has realized in close collaboration with me[10].

Figure 21.6: Picture from the Chinese Pavilion of the 2011 Venice Biennale.



9. For a brief account of this show and its relation to somaesthetics, see my discussion with curator Peng-Feng in *ArtPress*379, June 2011, Venice Biennale Supplement, 24-25.

10. An account of this experience (including images) can be found in Richard Shusterman, "A Philosopher in Darkness and in Light: Practical Somaesthetics and Photographic Art," in Anne-Marie Ninacs (ed.), *Lucidité. Vues de l'intérieur / Lucidity. Inward Views : Le Mois de la Photo à Montréal 2011* (Montreal: Le Mois de la Photo à Montréal, 2011).

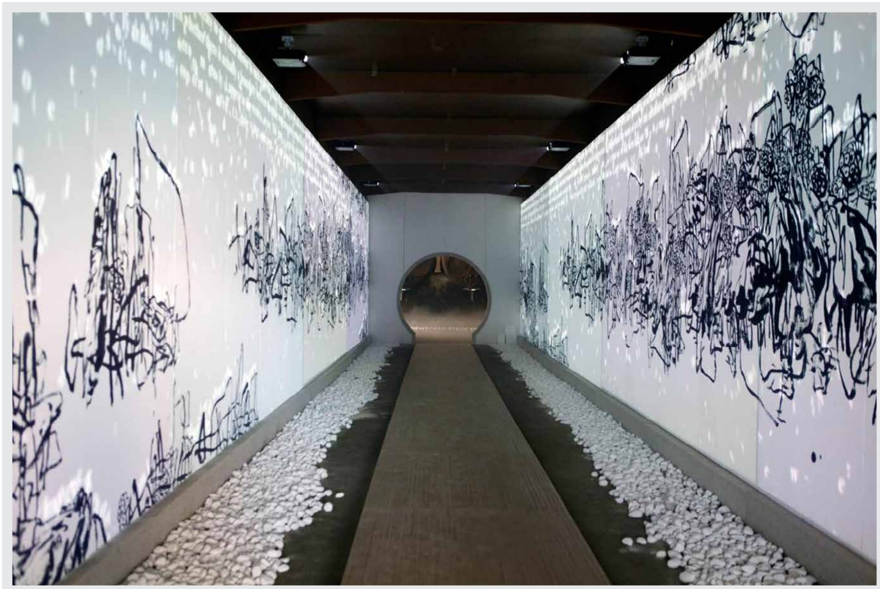


Figure 21.7: Picture from the Chinese Pavilion of the 2011 Venice Biennale.



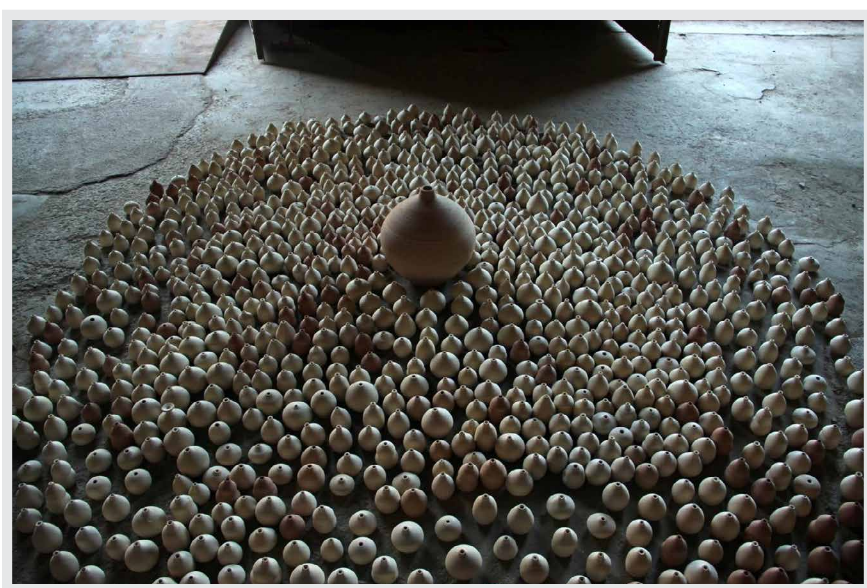


Figure 21.8: Picture from the Chinese Pavilion of the 2011 Venice Biennale.

Figure 21.9: Picture from the Chinese Pavilion of the 2011 Venice Biennale.

Figure 21.10: Picture from the Chinese Pavilion of the 2011 Venice Biennale.



Among political applications of somaesthetics, feminist interventions loom large. This should not be surprising since women are traditionally identified with body and thus negatively contrasted with what our culture deems to be the superior male principle of mind. As Shannon Sullivan uses somaesthetic ideas to critique the devalorization of bodily practices associated with women and to insist (through notions of somaesthetic teaching, caring, and dialogue) that working on the body is not a merely selfish, unsocial project, so Cressida Heyes deploys somaesthetics as a model enabling "political resistance to corporeal normalization" that subjugates women and men. Since race, like gender, is perceived through somatic appearance, racism provides another political issue in which somaesthetic strategies have been proposed both as explanations and as therapeutic remedies[11].

For me, the most surprising extension of somaesthetics has been in the arena of high-tech design, particularly with new information technologies. I did not expect this because the somaesthetic project was initially inspired by ancient ideas of the embodied philosophical life and by traditional Asian somatic practices such as yoga and *zazen* or contemporary Western counterparts (such as Alexander Technique or Feldenkrais Method) that preserve a similar organic character by not treating the body with electronic appliances. Although my work addressed the new media's challenge to embodiment, I did so mainly by arguing two major points: First, no technological invention of virtual reality will negate the body's centrality as the focus of affective, perceptual experience through which we experience and engage the world. Second, that cultivating better skills of body consciousness can provide us with enhanced powers of concentration to help us overcome problems of distraction and stress caused by the new media's superabundance of information and stimulation. But I made no effort to envisage positive ways that our newest technologies might reshape somatic experience. How, for example, could future developments in genetic engineering, nanotechnology, robotics, and experimental drugs yield significant changes in our somatic powers

11. See Shannon Sullivan, *Living Across and Through Skins: Transactional Bodies, Pragmatism, and Feminism* (Bloomington: University of Indiana Press, 2001); Cressida Heyes, "Somaesthetics for the Normalized Body," in *Self-Transformations: Foucault, Ethics, and Normalized Bodies* (Oxford: Oxford University Press, 2007), quotation from p. 124. David Granger, "Somaesthetics and Racism: Toward an Embodied Pedagogy of Difference," *The Journal of Aesthetic Education*, 44:3 (2010), 69-81. For an instructive overview of some of these political applications of somaesthetics, see Wojciech Malecki, *Embodying Pragmatism: Richard Shusterman's Philosophy and Literary Theory* (Frankfurt: Peter Lang, 2010), ch. 4.

either by changing the bodies which nature gives us or complementing them with prosthetic or chemical enhancements that dramatically augment the soma's perceptual, cognitive, and motor capabilities? How should somaesthetics prepare to deal with these changes and their corresponding new capacities for somatic self-cultivation, self-stylization, and social interaction?

Philosopher Jerrold Abrams has helped remedy this omission by exploring, in a speculative way, issues in what some might call posthuman somaesthetics because these issues involve significant alterations or enhancements to the traditional biological human soma[12]. Of course, the human soma is already a product of considerable evolution, and it seems plastic enough to absorb significant change and prosthetic devices without condemning us to being posthuman cyborgs. That we can sometimes be considered human cyborgs by having manufactured enhancements incorporated into our embodied selves (contact lenses, pace makers, false eyelashes, wigs...) seems more or less evident. Questions such as the possible limits of the human soma and whether or how should we should speak of nonhuman somas are interesting topics for somaesthetic analysis that I cannot properly address in this article. They depend not simply on the future of technology but also on the evolution of our conceptual schemes concerning the human and concerning the notion of soma[13]. But putting aside futuristic speculations, I will discuss some recent work relating somaesthetics to human-computer interaction research after the following section on the structure of somaesthetics

## 21.3 STRUCTURE OF SOMAESTHETICS

12. J. J. Abrams, "Pragmatism, Artificial Intelligence, and Posthuman Bioethics: Shusterman, Rorty, and Foucault." *Human Studies*, 27 (2004): 241-258; and "Shusterman and the Paradoxes of Posthuman Self-Styling," in Dorota Koczanowicz and Wojciech Malecki (eds.), *Shusterman's Pragmatism: Between Literature and Somaesthetics* (Amsterdam: Rodopi, 2012), 145-161.

13. Provisionally, I have suggested that higher animals do have somas but lack the status of selves or persons that humans typically have. For elaboration of these points, see Richard Shusterman, "Soma and Psyche," *Journal of Speculative Philosophy*, 24:3 (2010), 205-223.

Somaesthetics, as I conceive it, consists of three branches that sometimes overlap to some extent.

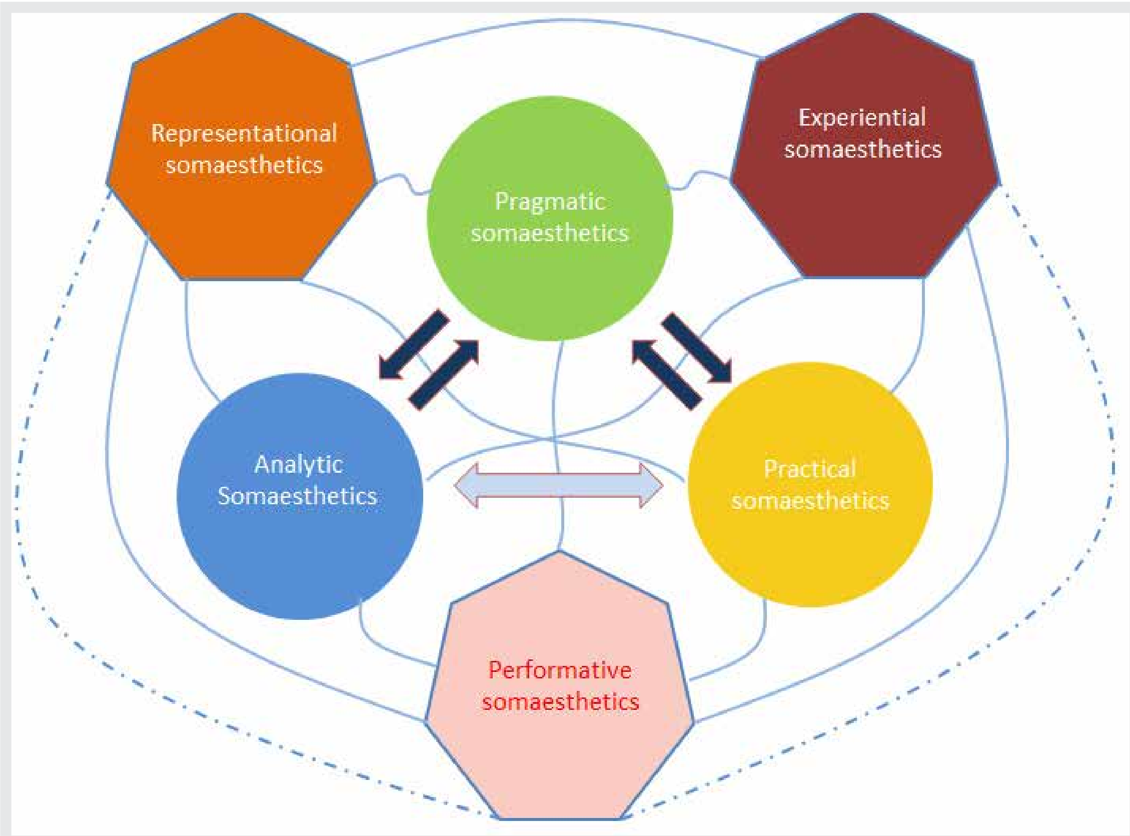



FIGURE 21.11: Diagram of the different branches and dimensions of somaesthetics and of their interrelations, designed by Richard Shusterman and Hyijin Lee.

*Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.*

### 21.3.1 Analytic Somaesthetics

The first, *analytic somaesthetics*, is an essentially descriptive and theoretical enterprise devoted to explaining the nature of our bodily perceptions and practices and their function in our knowledge and construction of the world. Besides the traditional topics in philosophy of mind, ontology, and epistemology that relate to the mind-body issue and the role of somatic factors in consciousness and action, analytic somaesthetics also includes the sort of genealogical, sociological, and cultural analyses of embodiment, including the body's role in sustaining social and political power. Such studies, most famously advanced by Simone de Beauvoir, Michel Foucault, and Pierre Bourdieu, show how the body is both shaped by power and employed as an instrument to maintain it — how bodily norms of health, skill, and beauty, and even our categories of sex and gender, are constructed to reflect and sustain social forces.

### 21.3.2 Pragmatic Somaesthetics

In contrast to analytic somaesthetics, whose logic is essentially descriptive, *pragmatic somaesthetics* has a distinctly normative, often prescriptive, character because it involves proposing specific methods of somatic improvement or engaging in their comparison, explanation, and critique. Since the viability of any proposed method will depend on certain facts about the body (whether ontological, physiological, or social), this pragmatic dimension presupposes the analytic dimension. However, it transcends analysis not only by evaluating the facts analysis describes but also by proposing methods to improve certain facts by remaking the body and the environing social habits and frameworks that shape it. A vast and complex array of pragmatic disciplines has been designed to improve our experience and use of our bodies: various diets, forms of grooming and decoration, martial and erotic arts, yoga, massage, aerobics, bodybuilding, calisthenics, and modern psychosomatic disciplines such as the Alexander Technique and the Feldenkrais Method.

These different methodologies of practices can be classified in different ways. We can distinguish between practices that are holistic or more atomistic. While the latter focus on individual body parts or surfaces — styling the hair, painting the nails, shortening the nose or enlarging the breasts through surgery — the former practices are emphatically oriented toward the whole body, indeed the entire person, as an integrated whole. Hatha yoga, t'ai chi ch'uan and Feldenkrais Method, for example, comprise systems of integrated somatic postures and movements to develop the harmonious functioning and energy of the body as a unified whole. Penetrating beneath skin surfaces and muscle fiber to realign our bones and better organize the neural pathways through which we move, feel, and think, these practices insist that improved somatic harmony is both a contributory instrument and a beneficial by-product of heightened mental awareness and psychic balance. Such disciplines refuse to divide body from mind in seeking the enlightened betterment of the soma (or body-mind) of the whole person.

Somatic practices can also be classified in terms of being directed primarily at the individual practitioner herself, or instead primarily at others. A massage therapist or a surgeon standardly works on others, but in doing t'ai chi ch'uan or bodybuilding, one is working more on one's own body. The distinction between self-directed and other-directed somatic practices cannot be rigidly exclusive since many practices are both. Applying cosmetic makeup is frequently done to oneself and to others; and erotic arts display a simultaneous interest in both one's own experiential pleasures and one's partner's by maneuvering the bodies of both self and other. Moreover, just as self-directed disciplines (like dieting or bodybuilding) often seem motivated by a desire to please others, so other-directed practices like massage may have their own self-oriented pleasures.

Despite these complexities (which stem in part from the deep interdependence of self and other), the distinction between self-directed and other-directed body disciplines is useful for resisting the common presumption that to focus on the body implies a retreat from the social. That presumption is surely wrong because not only is the body shaped by the social; it also contributes to the social. We can share our bodies and bodily pleasures as much as we share our minds, and they can be as public as our thoughts. Our bodies are visible social markers of our values, affiliations, and tastes. Somatic self-stylization generates an enormous commercial market that feeds the cosmetic, fashion, dieting, exercise, and plastic surgery industries, along with the advertising industry that supports them by stimulating our desire to stylize ourselves somatically. This desire typically takes the paradoxical form of wanting to fit in yet also stand out as distinctive. In other words, self-styling involves conforming in some way to the norms of some social taste group (even if it be a subculture that resists mainstream taste) yet not allowing such conformity to group style to preclude one's own individual expression[14].

Moreover, it is crucial to remember that caring for one's own body is essential to caring properly for others, since all helpful action requires bodily means. That is why the Confucian commandment to respect and care for one's parents carries with it the command to preserve one's own bodily welfare. My professional work as a Feldenkrais practitioner has taught me how important it is to pay careful attention to one's own somatic state in order to pay proper attention to one's client. When I give a Feldenkrais lesson of Functional Integration, I have to be aware of my own body positioning and breathing, the tension in my hands and other body parts, and the quality of contact my feet have with the floor in order to be in the best condition to gauge correctly the client's body tension, muscle tonus, and ease of movement[15]. I need to make myself somatically comfortable so as not to be distracted by my own body tensions and in order to communicate the right message to the client. Otherwise, I will be passing my feelings of somatic tension and unease to the client when I touch him. And since one often fails to realize when and why one is in a mild state of somatic discomfort, part of the Feldenkrais training is devoted to teaching one how to discern such states and distinguish their causes.

Figure 21.12: A somaesthetics body-scan lesson taught by Richard Shusterman, photographed by Hyijin Lee.

14. For more detailed discussion of somatic style and its somaesthetic modalities, see Richard Shusterman, "Somatic Style," *Journal of Aesthetics and Art Criticism* 69:2 (2011), 147-159.

15. Feldenkrais Method deploys an educational rather than therapeutic-pathological model. Practitioners thus work with clients who are treated as "students" rather than "patients", and we speak of our work as giving "lessons" rather than "therapy sessions". I describe the Feldenkrais Method in greater detail in chapter 8 of *Performing Live*. Functional Integration is only one of the two central modes of the Method, the other being Awareness Through Movement. The latter is best described in Feldenkrais's introductory text, *Awareness Through Movement* (New York: Harper and Row, 1972).





Figure 21.13 A-B: A Feldenkrais movement posture; a somaesthetics workshop lecture photographed by Hyijin Lee.

Clearer awareness of one's somatic reactions can also improve one's behavior toward others in much wider social and political contexts. Much ethnic and racial hostility is not the product of logical thought but of deep prejudices that are somatically marked in terms of vague uncomfortable feelings and are thus engrained beneath the level of explicit consciousness. Such prejudices and feelings thus resist correction by mere discursive arguments for tolerance, which can be accepted on the rational level without changing the visceral grip of the prejudice. We often deny that we have such prejudices because we do not realize that we feel them, and the first step in controlling or expunging them is to develop the somatic awareness to recognize them in ourselves[16].

Somatic disciplines can further be classified as to whether their major orientation is toward external appearance or inner experience. Representational somaesthetics (such as cosmetics) is concerned more with the body's surface forms, while experiential disciplines (such as yoga) aim more at making us feel better in both senses of that ambiguous phrase: to make the quality of our somatic experience more satisfying and also to make it more acutely perceptive. The distinction between representational and experiential somaesthetics is one of dominant tendency rather than a rigid dichotomy. Most somatic practices have both representational and experiential dimensions (and rewards), because there is a basic complementarity of representation and experience, outer and inner. How we look influences how we feel, and vice versa. Practices such as dieting or bodybuilding that are initially pursued for representational ends often produce inner feelings that are then sought for their own experiential sake. Just as somatic disciplines of inner experience often use representational cues (such as focusing one's attention on a body part or using imaginative visualizations), so representational disciplines such as bodybuilding use experiential clues to serve their ends of external form, helping to distinguish, for example, the kind of pain that builds muscle from the pain that indicates injury.

16. I elaborate this argument more fully in *Body Consciousness*.

A third category of pragmatic somaesthetics could be distinguished for disciplines that focus primarily on building strength, health, or skill and that would include practices such as weightlifting, athletics, and martial arts. This category could be called "performative somaesthetics." But to the extent that these disciplines aim either at the external exhibition of performance or at one's inner feeling of power and skill, they might be associated with or assimilated into the representational or experiential categories.

### 21.3.3 Practical Somaesthetics

The different methodologies of pragmatic somaesthetics need also to be distinguished from actual somatic practice, which I construe in more robust terms than the mere writing and reading of body-related texts, even those outlining pragmatic methods. Thus, besides analytic and pragmatic branches of somaesthetics, there is a further branch — *practical somaesthetics* —which involves actually engaging in programs of disciplined, reflective, corporeal practice aimed at somatic self-improvement (whether representational, experiential, or performative). This dimension of not just saying but of physically doing seems sadly neglected by contemporary accounts of the body in philosophy and other humanities disciplines, though it has often been crucial to the idea of the philosophical life, and it is essential to the idea of somaesthetics as integrating both theory and practice. I therefore teach workshops on practical somaesthetics to convey this practical dimension in a more embodied way than merely verbal insistence.

FIGURE 21.14: A video clip extract from a somaesthetics workshop for dancers and choreographers (in French), filmed by Damien Marteau.

## 21.4 SOMAESTHETICS AND HUMAN-COMPUTER INTERACTION

Researchers and practitioners in the field of Human-Computer Interaction have been increasingly engaging with somaesthetics in their work, particularly in the field of HCI design. Though their interest initially surprised me, I should have expected it because the soma is central to everything that we do. Not only does it serve as the basic tool through which we perceive the world and deploy all further tools (including computers) but the soma is also our most intimate example of an interactive system of interdependent (but also to some extent autonomously functioning) interacting subsystems with extensive, complex, subtle, yet amazingly rapid and efficient feedback loops. In plotting how computers interact with human systems, the multifaceted somatic complex system should be at the core. At present somaesthetic-related HCI design research includes both theoretical models and more concrete productions. My discussion will choose an example from each.

One promising theoretical effort (developed by Youn-Kyung Lim and her colleagues) proposes a model that integrates the basic sensory and affective experience of the computer user together with the physical properties of the tools deployed in computer interaction and then explains how these and other factors produce higher, emergent qualities of interactive aesthetic Gestalt that belong to the overall interactive situation or experience[17]. In this model, somaesthetics not only provides recognition that somatic feelings form an important part of user experience in the interaction but it also offers methods to heighten the user's (including the designer's) consciousness of those somatic feelings so their feedback into the interactive Gestalt can be more effective (and, with the designer, more effectively available for improving the design).

At this stage in somaesthetics research, we have only been concerned with somatic feelings of human bodies and thus with only one side of the HCI interaction. But, in principle, it may be possible to consider the somaesthetics of non-human somas, including computer bodies. If these are complex, sensitive, and responsive enough in their perceptions and reactions, perhaps we can eventually speak, in some way, of their somatic experience in the interaction. Perhaps we can speak of their own bodily experience of (or physical response to) rough or clumsy handling or smoothly flowing use, even if there is no good way of attributing to these computers human-like conscious feelings. This idea of computer or robotic somaesthetics may seem wildly futuristic, and it seems more promising for now to focus somaesthetic research on human somatic experience. But it would be wrong to preclude in principle that somaesthetics could be developed beyond the human soma to make its contribution to HCI research still richer by dealing with both sorts of bodies — organic and mechanical, particularly since the human soma is increasingly lived through mechanical enhancements, including such traditional ones as eyeglasses and hearing aids.

17. Youn-Kyung Lim, Erik Stolterman et al. "Interaction Gestalt and the Design of Aesthetic Interactions," ACM, New York, NY 2007, 239-254. For more recent applications of somaesthetics to the theory or methodology of design, see, for example, Petra Sundström, Elsa Vaara, Jordi Solsona, Niklas Wirström, Marcus Lundén, Jarmo Laaksolhati, Annika Waern, and Kristina Höök, "Experiential Artifacts as a Design Method for Somaesthetic Service Development," RDURP'11, Sept. 18,2011, Beijing, China (ACM 978-1-4503-0931-8/11/09).

With respect to the lived integration of human and computer bodies I can bring an intriguing example of a practical application of somaesthetics to HCI research. Thecla Schiphorst "argues for the value of exploring design strategies that employ a somaesthetic approach" not only through theoretical texts but also by fashioning a series of interactive, networked artworks based on sensory interactions involving touch and breathing. Some of these works are "interactive wearable art" in which the garments react not only to the wearer's organic movements and breathing but also to tactile or breathing inputs from other participants interacting through computer networks, which includes iPhone inputs. Other works are soft computerized bodies that react to human touch (but also to their own movement) by responses of vibration, light, and sound which they can communicate wirelessly (to enlarge the network of interactive response) to other such bodies in the network[18].

## 21.5 PRACTICAL APPLICATIONS

The practical applications of somaesthetics are as wide-ranging and diverse as the uses of the soma in our lives, for it is the core medium of all our perception, cognition, and action. Its applications to the arts, to health and fitness, to socio-political issues such as racism, sexism, and ethnic hatred, and to education have been discussed extensively elsewhere[19]. I confine my remarks here to applications in product design, with special attention to HCI. But I believe that the design of any product (chair, cereal box, car, or cell phone) could beneficially employ somaesthetic principles. Somaesthetics is not ergonomics. One problem with ergonomics

18. Thecla Schiphorst, "soft(n): Toward a Somaesthetics of Touch," CHI 2009, ACM, New York, NY, pp. 2427-2438; quotation p. 2427; and Thecla Schiphorst, Jinsil Seo, and Norman Jaffe, "Exploring Touch and Breath in Networked Wearable Installation Design," MM 2010, ACM, New York, NY, 1399-1400; quotation 1399. See also the application of somaesthetics to computer games in H. S. Nielsen, "The Computer Game as a Somatic Experience," *Eladamus. Journal for Computer Game Culture*, 4:1 (2010), 25-40.

19. Some of my more important papers discussing these topics are collected and revised in my new book, *Thinking Through the Body: Essays in Somaesthetics*, forthcoming from Cambridge University Press.

is its interactive deficiency; it does not really involve the subject's input in a serious way, and certainly not in a continuous way. In contrast, somaesthetic design does involve user input because it is essentially structured on body consciousness, how the user's body actually feels in the relevant actions performed and not simply how it is anatomically structured in general or in the abstract. Somaesthetics works not only with the level of explicit consciousness of bodily feelings but with reflective somatic consciousness, that is with how awareness of the somatic feeling modifies that feeling. In this way, somaesthetics can provide for more effective, richer feedback loops in interaction. Moreover, somaesthetic recognition that our bodily perceptions and feelings are transmodal can help designers avoid the mistake of not taking into account sensory integration but rather base their interactive design by considering individual senses in isolation and treating them as separate capacities (often to be ranked and hierarchized).

Somaesthetics appreciation of the particularities of individual body consciousness offers a further advantage for interactive design. Interactive products are not necessarily sensitive to individual tendencies so that each person must try to conform to the standards set by designers, which are often arbitrary and based on visual form. Somaesthetics can help design become more effective and pleasurable to the end-user not only by taking into account of bodily feelings appropriate to the product but also personalizing them based on the computation of an individual's preferences, habits, and performances. An example would be a touch screen that automatically calibrates the weight of an individual's touch, so that using the product would feel easier and more agreeable rather than frustrating, because it would be more responsive to the individual user. This greater responsiveness, if developed through changing input, could make the product interaction also more meaningful to the user, providing an input on his or her current state. Another example might be a hand-held object like a cell phone; how does the object feel in the hand. The emphasis on lightness in design can be exaggerated. Sometimes things with a greater weight feel better in the hand and are more effectively held and used.

The reasons for a heavier object feeling better to us can be very complex. They include not only sensory reasons and habitual expectations of certain weights but also psychological associations, for example the association of weight with sturdiness, durability, and thus reliability. The Danish audio-visual company Bang and Olufsen actually add unnecessary metal to their hand-held remote control products to give their consumers confidence that these products are well-made and sufficiently rugged to endure hard use.


Figure 21.15: The B&O remote control — heavy seems to mean sturdy.

More generally, designers could improve their design skills by becoming more aware of how they use themselves and how they feel when using particular products rather than merely thinking of how the product is conventionally used. A comparative study of how different shapes of cup handles affect one's feelings towards drinking, for example, could be used to improve cup design. Does the handle make your forefinger grip tightly, and how does this affect the rest of your arm, the rest of your body, and by extension, the feeling of drinking?

## 21.6 FUTURE DIRECTIONS

Rooted in ancient philosophical ideas and body disciplines that have been reconceptualized through contemporary pragmatist philosophy, somaesthetics is firmly grounded in philosophy, history, and theory, but its future directions, I hope, will be increasingly interdisciplinary and practical. It is a vast and extremely diverse research project that can welcome a wide variety of researchers. The most profitable interdisciplinary engagement I envisage for somaesthetics would not be a mechanical application of somaesthetic principles derived from philosophical speculation and then applied to another field such as health, design, art, and so on. A more rewarding future is for interdisciplinary teams to work together on somaesthetic questions in which experts in somaesthetic theory, disciplines of body consciousness, and other disciplinary fields interested in applying somaesthetic ideas would dialogue and experiment together. It is hard to combat disciplinary inertia due to the professionalization of knowledge in terms of compartmentalized disciplinary structures and specialties, each with their restricted vocabularies and restricting gatekeepers. One way of tackling that problem is through a practical workshop setting in which theoretical ideas not only can be exemplified and tested in real life somatic actions but can also be refined and new ideas generated through experiential input that is filtered and elaborated through transdisciplinary communication but rooted in a common experiential process structured by the workshop protocol. As disciplines such as HCI show greater interest in somaesthetics, I envisage working with IT and design experts in developing criteria for somatic profiles of use, comfort, ease, and pleasure that could be employed in interactive design. The field of robotics is another HCI area to which somaesthetics could contribute by analyzing how people feel in interacting with other bodies (both human and mechanical) in space. Such studies could teach us what robotic movements (and not simply what robotic shapes) are friendlier for human interaction.

## 21.7 WHERE TO LEARN MORE

Shusterman, Richard (2012): *Thinking through the Body: Essays in Somaesthetics*. Cambridge University Press

In addition, there is a somaesthetics bibliography of the author and a somaesthetics bibliography of other authors.

## 21.9 COMMENTARY BY JEFFREY BARDZELL

How to cite this commentary in your report

### Jeffrey Bardzell

*© Jeffrey Bardzell*

Jeffrey Bardzell is an Associate Professor of HCI/Design and new media at the School of Informatics in Indiana University - Bloomington. With a Ph.D. in Comparative Literature and Minor in Philosophy, Bardzell brings a humanist perspective to HCI and is known for developing a theory of interaction criticism. His other HCI specialties include aesthetic interaction, user experience design,...

Jeffrey Bardzell
Jeffrey Bardzell is a member of The Interaction Design Foundation

I begin my commentary by expressing my excitement to see an important philosopher directly engaging the HCI community. Up till now, philosophy's participation in HCI has been mediated by HCI researchers' interpretations and often (frankly) dumbed-down introductions to philosophical concepts. So this is a special moment for us, but it also introduces a practical challenge, and that is that because Shusterman is outside of our community, he can't fully participate in our "language games," which, in the case of HCI refers to the ways we frame and legitimate research, the sorts of questions we ask, the stock of examples we assume everyone knows, the history of the field we all think we hold in common, etc.

For a relationship between somaesthetics and HCI to prosper, linkages will need to be developed. Doing so is a book or at least a lengthy journal article, and I have here only a commentary to work with. So I will pursue the more modest goal of sketching out a position for somaesthetics in HCI, with the hope that others (from HCI, from philosophy) will join those of us who are already exploring those linkages.

My thesis in this commentary is that somaesthetics occupies a unique theoretical position in our field, able to connect pragmatist approaches to HCI (design theory, experience design) and embodied approaches to HCI (affective computing; mobile, pervasive, and ubiquitous computing). This has implications both for users, and in particular, **norms for serving users in the deepest and most important ways possible**; and also for interaction designers, and in particular, **the cultivation of the professional self as an expert subject**.

My argument will first sketch out recent calls for a more designerly HCI, exploring what those calls really mean, and specifying the challenges this community faces in responding to them; and second it will explore somaesthetics (and related pragmatist traditions) as a collection of useful concepts that has practical, pedagogical, and normative implications for a somaesthetic HCI. I will conclude by reflecting on some limitations of somaesthetics and HCI, which could guide future work.

### 21.9.1 The rise of design as an input for HCI

As is well documented elsewhere, since the 1990s there have been increasing calls for a more designerly approach to HCI, calls which have picked up steam since 2000. As computing has moved from the workplace and into everyday life, e.g., with the rise of mobile phones, pervasive computing, and the wild spread of digitally enabled consumer electronics, from DVRs to programmable sex toys, the distinctions between product design, service design, communications design, interaction design, and the emerging area of experience design have never been blurrier. Today design is a thriving subcommunity of HCI, with its own official subcommunity within CHI, a highly successful biannual conference in DIS, and ongoing calls for even more design contributions to the field. For example, McCarthy & Wright (2004) observe that

> "*there has been a perceptible shift in nomenclature toward Interaction Design or User Experience Design when referring to relationships between people interactive technologies. This reflects a broadening of focus from computers to a wide range of interactive technologies and from work-related tasks to lived experience.*"
>
> -- *McCarthy and Wright (2004): p. 3*

This turning towards design is often couched in critiques of the limits of social scientific approaches that have dominated HCI for decades. Greenberg & Buxton (2008) write,

> "*One way to recast this is to propose that the subjective arguments, opinions, and reflections of experts [e.g., trained designers] should be considered just as legitimate as results derived from our more objective methods.... Another way to recast this is that CHI's bias towards objective vs. subjective methods means it is stressing scientific contribution at the expense of design and engineering innovations.... The net result is that we eliminate ideas too early, we consider far too few ideas at all, we converge on that which we can measure, which is almost always that which we are already familiar with. Our work degrades into a refinement of the known rather than innovation along new trajectories* "
>
> -- *Greenberg and Buxton (2008): pp. 114-15*

There is a lot going on in this quote, including a critique of an HCI dogma that privileges objective measures over other legitimate forms of design knowledge-making, a prioritization of the social sciences over design and engineering innovation, and above all the practical consequences that we are too narrow to innovate. Restating this argument more simply, Greenberg and Buxton imply that innovation depends on subjective expertise and is hard to get at using the objective methods valorized by the HCI community.

Kuutti (2009) agrees: tracing the evolution of HCI as an academic-industry collaboration to deal with the explosion in demand for usable software during the rise of the PC in the 1980s, Kuutti describes how *usability* became the primary practical achievement to respond to this problem, with the unfortunate side-effect that "this also meant a certain intellectual impoverishment: calls to discuss about HCI theoretical foundations lost the audience, when the somewhat a-theoretical (and originally sometimes even anti-theoretical) usability movement took over" (Kuutti, 2009, p. 7); indeed this anti-theoretic stance is alive and well today, as can be seen in Tractinsky's account of visual aesthetics in HCI and the portion of my commentary devoted to confronting his anti-theory position.

So the lack of design thinking inhibits innovation, and it is characterized by impoverished theory and a dogmatic fetish for objectivity. What then *is* a designerly approach to HCI? What sorts of theories, methodological strategies, and goals add up to a designerly HCI? A few choice quotes can at least offer some insights:

> "*Design education has also not been heavily interested in training students in detailed methods how to do design, but more educating such personalities who can filter and crystallize cultural influences into effective and meaning-laden forms. So the development of personal judgment what is good design what not has always been one of the goals in [design] education. This also means a certain individualism; it is assumed that a design brief interpreted by two designers will lead to two different designs.* "
>
> -- *Kuutti, 2009: p. 3*

Design pedagogy is about "educating such personalities that can filter and crystallize cultural influences into effective and meaning-laden forms." This characterization is a far cry from the traditional rhetoric of user-centered design. Whereas user-centered design positions the designer in an almost passive position of discovering existing needs using scientific methods and then designing around and for what is discovered in that activity, traditional design activates the designer as a perceptive, insightful, and imaginative meaning-maker, an ability that is individualistic to a certain degree and dependent on judgment rather than data, and offers a radically different view of the foundations of a design problematic.

Cockton offers a more specific account of the sorts of things that designers do as a part of their profession:

> "*The gaps [in traditional HCI] include a lack of: ways to track and reflect on design purpose; theoretical receptiveness; underpinning transdisciplinary theory to scope such sensitivities; well thought through approaches for non-work settings; and ways to maintain and compare a diverse range of alternative design means....*"
>
> -- *Cockton, 2008: p. 2482*

This is a different sort of list than one might expect from a traditional textbook on HCI: reflection, theory, a foregrounding of purpose, transdisciplinarity, thoughtfulness, exploration of alternative design means all characterize a designerly approach. Again, the role of a subjective expertise is unmistakeable in this list. It is quite easy to imagine how one designer could be superior at reflection, use of theory, creativity with alternative means, and so on than another. In contrast, user-centered design methods seem to presume that a designer's strength is the quality of the data informing decisions, not anything internal to the designer herself.

My third quote comes from Greenberg & Buxton again, who offer a different sort of list and yet one that is fundamentally compatible with the vision of design that is emerging here:

*"there are many other appropriate ways [besides usability evaluation] to validate one's work. Examples include a design rationale, a vision of what could be, expected scenarios of use, reflections, case studies, participatory critique, and so on. At a minimum, authors should critique the design.... Early evaluation is usually done through the Design Critique (or 'Crit'). The designer presents the artifact to the group (typically a mix of senior and junior people), and explains why the design unfolded the way it has. Members of the group respond: by articulating what they like and dislike about the idea, by challenging the designer's assumptions through a series of probes and questions, and by offering suggestions of ways to improve the design. This is a reflective and highly interactive process: constructive criticisms and probing demands that designer and criticizers alike develop and share a deep understanding of the design idea and how it interacts within its context of use."*

*-- Greenberg and Buxton: p. 118-119*

Guiding all of this seems to be a holistic *interpretation* of what the design will be, accompanied by a rationale, and vetted by an intensely iterative and ongoing critical process involving stakeholders and other designers. This situates design in a dialogic and argumentative tradition. What is being argued for and against is the designer's *particular framing* of the problem and *speculative vision* of its solution. Again, the subjective expertise of the designer—as an active meaning-maker and speculative reasoner—is the foundation of the whole activity. Others are brought in on similar terms: their own ability to interpret, frame, re-frame, and speculate determines the quality of critical feedback that they can provide.

In short, design professionals require a cultivated ability to read socio-cultural signs and trends; a creative and reasoned ability to explore alternative futures; a verbal ability to articulate these activities; a receptiveness to alternative framings and a willingness to explore highly variable alternative directions; and above all a personal identity or coherence that holds all of these moving parts together through a given process. Much more is personally demanded of designers than is personally demanded of traditional usability engineers.

So how to we get there?

There are several answers to this question. One is that a small number of individual designers cannot achieve critical mass: we need a design culture in HCI (Nelson & Stolterman, 2003). Creating such a culture has implications both for how interaction designers are trained and also for how the community legitimates certain knowledge practices. But however that happens, one thing is for certain: theory is going to be in the middle of it. Theory has historically been central to design and the humanities, inasmuch as each is concerned with the insightful and imaginative understanding of culture. Inasmuch as HCI now wants such accomplishments to be part of its discipline, HCI will have to get over its aversion to theory and fetish for methods. And this is where a philosophical program such as somaesthetics enters the picture.

### 21.9.2 Somaesthetics in/for HCI

Shusterman has offered here an accessible introduction to somaesthetics, effectively condensing and synthesizing earlier works, including "Somaesthetics: A Disciplinary Proposal," which appeared as Chapter 10 in the second edition of *Pragmatist Aesthetics: Living Beauty, Rethinking Art* (2000) and his more recent book *Body Consciousness* (2008). I reference these books partly because they much more context for somaesthetics and Shusterman's whole project. The first chapter of *Pragmatist Aesthetics* carefully situates somaesthetics within the context of Dewey's pragmatism and in opposition to analytic aesthetics. This context translates to some core positions.

One core position is a formulation of somaesthetics as comprising **a holistic and even organic perspective on life, work, and the self**. Shusterman's holism is evident in many places, not least in his efforts to undermine rigid distinctions between aesthetics and ethics, as when he writes, "aesthetic considerations are or should be crucial and ultimately perhaps paramount in determining how we choose to lead or shape our lives and how we assess what a good life is" (Shusterman, 2000, p. 237). HCI has long debated about the relative values of function versus aesthetics, with traditional HCI siding with function; but Shusterman follows Dewey in rejecting the distinction and seeing aesthetics as a holistic and inclusive term that encompasses the categories traditionally subsumed under function and form.

Dewey's notion of an aesthetic life is not mere hedonism, but rather a sophisticated understanding of how the human as an organism purposefully and successfully copes with its environment (Berstein, 1971). Shusterman follows Dewey in rejecting divisions, distinctions, formalisms, and hierarchies. This is not merely an abstract chin-scratching philosophical position, but rather a practical position that has serious methodological implications: it basically rejects atomism or scientific reductionism of experience (Bernstein, 1971). Applied to HCI, such a view would reject the ways that affect researchers decompose affect into mood and emotion, emotion into positive and negative valence, positive valence into *n* number of positive emotions, and each of those into *d* degrees of intensity. Rather, a Deweyan view would construe emotion as a part of human's purposive future-oriented disposition to the world, helping the organism orient itself in the best possible way.

Described holistically, somaesthetics is "**a life-improving cognitive discipline** that extends far beyond questions of beauty and fine arts and **that involves both theory and practical exercise**," which seeks "to end the neglect of the body that [was] disastrously introduced into aesthetics," with the ultimate goal to "contribute significantly to … **an art of living**" (Shusterman, 2000, pp. 266-67). That's a nice sounding agenda, but what does all that mean?

Somaesthetics, like any other philosophical position, can be characterized as comprising a system of relating concepts. I began with the context and overall view, because I don't want an analysis of somaesthetic concepts applied to HCI to lose sight of what somaesthetics is supposed to *do*, which is to help us lead or shape our lives and to recognize what a good life is. I will consider this in two different directions, both of which are central to HCI:

- The training of interaction designers
- Normative criteria for user experience

#### 21.9.2.1 Aesthetic perception, somatic training, and the interaction designer

In this section I want to argue that two of Shusterman's key concepts, aesthetic perception and somatic training, contribute to both (a) a substantive epistemological account of the designer as expert subject and also (b) a useful set of norms to orient professionals cultivating designerly ways of doing (with implications for interaction design pedagogy as well).

##### 21.9.2.1.1 AESTHETIC PERCEPTION

Among the most common views of aesthetic perception in HCI is the stimulus-response model. On this view, an object in the world, such as an interface, acts as a stimulus to the human cognitive system, which responds to it, e.g., by perceiving it, storing it in memory, understanding it, deciding to act based on it, etc. In HCI, for example, physiological data, such as breath rate and skin conductance, is collected as ways of measuring a person's response to an input. Important advances to the field have been made with this model, and I certainly don't want to suggest that it is somehow "wrong."

But the stimulus-response model has epistemological limitations, and these have implications for interaction design professionals. The key limitation is that such a model assumes the existence of certain interpretative skill in the first place. It's not always obvious to us, but understanding is a deeply cultural and learned ability. For instance, when we go to a museum or historical site, a docent not only relates historical backgrounds and contexts, but more importantly *tells us what to look at*. But if it is right before our eyes all along (i.e., a visual stimulus), then why don't we respond the right way (i.e., with appreciation)? Why is it that a professional and amateur photographer standing side-by-side looking at the same thing will take very different pictures of it? How can a professional designer look at a given design material and come up with surprising and expressive new forms, where others simply rehash existing forms? Docents, accomplished photographers, and designers see and understand in richer ways than others do, and this is fundamental to their professional abilities.

Shusterman offers an explanation for these abilities by paraphrasing the work of 18th century philosopher Baumgarten, who first coined the term "aesthetics." For Baumgarten, in Shusterman's paraphrase,

> "*The end of aesthetics … is the perfection of sensory cognition as such, this implying beauty"…. Baumgarten insists especially on "keenness of sensation," "imaginative capacity," "penetrating insight," "good memory," "poetic disposition," "good taste," foresight," and "expressive talent.*"

-- *Shusterman, 2000*: *pp. 264-65*

In the stimulus-response model, and in most empirical science itself, such qualities have no meaningful place. Visual stimuli, and by extension empirical data, are seen to speak for themselves: the experiencing subject has only to perceive them to understand them. So stimulus-response is basically passive, and the model doesn't differentiate among responders. Much UX research in this tradition assumes that research subjects are fundamentally interchangeable and simply seeks to average their physiological or reported emotional responses (e.g., using Likert scales).

But what Shusterman wants to emphasize is—and here he is leveraging phenomenological hermeneutics and reader-response theory as well as pragmatism—that **understanding and hermeneutic skill must also exist before perception**. Though we often speak commonsensically as if object, lightwaves, visual perception, mental image, understanding, judgment, and decisionmaking all happened in a causal linear sequence, in fact it cannot characterize what actually happens. Meaning-making is an active process; meaning is not a form stamped in our cognition like a seal ring to wax.

Some people can perceive more keenly than others; some have more penetrating insights, some have a greater imaginative capacity. Importantly, these are not static "faculties" that we are born with, but rather "sensory cognition" comprises cultivatable abilities or habits that we practice and can improve over a lifetime. As Dewey writes of intelligence, it is not "the faculty of intellect honored in textbooks" [but rather is] "the sum total impulses, habits, emotions, records, and discoveries which forecast what is desirable and undersirable in future possibilities, and which contrive ingeniously on behalf of imagined good" (Dewey, cited in Bernstein, 1971, p. 211).

Somaesthetics is substantially responsive to the calls for a design sensibility in HCI because it offers an epistemological account of what such a sensibility actually is: a sensitive, imaginative, penetrating, tasteful, poetic, and expressive habit or disposition to design problems, materials, processes, opportunities, and situations. But how does one achieve such habit or disposition?

#### 21.9.2.1.2 SOMATIC TRAINING

One of the signature pieces of Shusterman's somaesthetics is his call for somatic training: if it's the case that the body is the "tool of tools," then philosophers need to get out of their armchairs and cultivate their own somatic competencies. He defines this as "actually engaging in programs of disciplined, reflective, corporeal practice aimed at somatic self-improvement" (21.3.2).

To understand what he means by this, we might consider the practices that he uses to exemplify this: "diverse diets, body piercing and scarification, forms of dance and martial arts, yoga, massage, aerobics, bodybuilding, various erotic arts (including consensual sadomasochism), and such modern psychosomatic therapies as the Alexander Technique, the Feldenkrais Method, Bioenergetics, Rolfing, etc." (Shusterman, 2000, p. 272). Common to each of these is a long-term commitment to body refinement. This is not a question of mastering the body by feeding and exercising it according to recommendations from the health sciences; rather, this is a practice of self-stylization for which the body is the locus of one's individual stylized identity. The distinction I'm drawing here is akin to shaving one's head versus becoming a "skinhead." One is a mere physical description of a change to the body, whereas the other entails the same physical change but in such a way that it is inscribed in and defining a symbolic identity for the person doing it.

Thus, for Shusterman, the cultivation of somatic sensibility is an outcome of corporeal training. And surely he is right about this. I trained myself to skate well enough to join an amateur hockey league, and I also enjoy watching ice hockey. It is certainly the case that my appreciation for the watching the sport is partly based on a *somatic* understanding of the sheer skill of good players: people who have themselves tried to skate backwards while turning and accelerating and also while controlling a puck and keeping it from an opposing player can almost physically appreciate the somatic aesthetics of such movements as they are displayed gracefully and effortlessly in a professional game. Surely people who go to the ballet, many of whom have had some dance training themselves, also appreciate somatically, almost "feeling" in their own bodies the somatic near-perfection of a professional dancer's alignments, body angles, and turnout. Indeed, how many somatic spectacles (e.g., professional sports, dance, rock concerts, etc.) do we enjoy that we *haven't* tried out in one form or another, whether it's backyard football, an *attitude en pointe* in front of the mirror, air guitars, or temporary tattoos?

Somatic training is also a part of HCI. We don't just talk about designs: we sketch them, prototype them, try them out, put them in people's hands and homes and watch what they do. These are all somatic exercises, and all of them require considerable training before anyone becomes good at them. Even among HCI researchers, the rising interest in critical design (e.g., Dunne, 2006; Dunne & Raby, 2001) and research through design (Zimmerman et al., 2007) extends this trend: such research uses design methodologies not in the hopes of creating new commercial products, but in order to generate knowledge and theory. Critical design is not easy to do, and seems to be an activity that requires iteration, practice, and training (Bardzell et al., 2012).

The other obvious area of HCI that involves somatic training is pedagogy. In our HCI/Design Master's program at Indiana University, students work in groups and individually to generate sketches, prototypes, workbooks, and portfolios within a studio culture in which they are frequently presenting their work for critique by faculty and peers. I don't think we're unique in that: such pedagogy is a part of our field. Additionally, a majority of our students' Master's projects involve a domain with which—external to their participation in our program—they already have somatic training: we've seen projects building on prior experience in political activism, hardcore World of Warcraft play, fashion illustration, senior health care provision, and film production, among others.

Somatic training is substantially responsive to the call for a design sensibility because it relates design processes and practices to the underlying epistemology of an expert subject. Design processes are a form of somatic training: they entail disciplined embodied practices, and these practices eventually heighten perceptual and expressive sensitivities towards human needs, visual forms, problem reframings, socio-cultural meanings. It is by such mechanisms that designers train to become the kind of people who can, re-quoting Kuutti, "filter and crystallize cultural influences into effective and meaning-laden forms."

#### 21.9.2.2 Somaesthetic Experiences

Much of my commentary thus far has focused on the suitability of somaesthetics as a theory that offers a rich and useful account for training the specialized sensibilities expected of design professionals. However, neither Shusterman nor the other pragmatists were seeking to support specifically the design profession: the pragmatist project is fundamentally geared to improving all human quality of life by reminding us that humans are organisms purposefully engaged in their lived environments, and not information processors or "disembodied ratiocinators" (in the memorable phrasing of Bannon & Bødker, 1991). And that means that somaesthetics also provides normative criteria for the design of aesthetic experiences for users.

That is, if we want to reframe UX away from usability and towards something more robustly aesthetic, then we need to replace existing UX goals with new ones. Traditional ones include low task completion time, low error rates, high learnability, etc. Kutti (2001) proposes three alternative criteria that seem in the spirit of Deweyan pragmatism:

- Users as learners
- Users as shapers of their environment
- Users as becoming something else by using a device or a system

These sound like good goals, but their abstractness creates a practical challenge for designers. Here somaesthetic and pragmatist theory can begin to unpack some of these concepts. From a pragmatist perspective, aesthetic interaction should contribute to some combination of the following user experience-abilities. I combine "experience-abilities" into a single concept, because pragmatists stress how skills emerge in and from experiences; it is only by being challenged—not dumbed down with ease of use, transparency, and simplicity—that skills and aesthetic pleasure emerge.

The following are normative goods valorized by pragmatism in general and somaesthetics in particular. That is, an aesthetic interaction is one that adheres or contributes to some critical mass of the following:

- The experience and cultivation of Baumgarten's "**perfection of sensory cognition**," that is:
    - keenness of sensation
    - imaginative capacity
    - penetrating insight
    - good memory
    - poetic disposition
    - good taste
    - foresight
    - expressive talent (from Shusterman, 2000, pp. 264-65)

- The expansion of people's range of **sympathetic identification with others** (Guignon & Hiley, 2003, p. 36, paraphrasing American pragmatist Richard Rorty)
- A reformation of *objects* not as external to and in opposition to indi-

vidual subjectivity, but rather seeing **objects as contiguous with human consciousness**, that is, seeing objects as human "activity in an objectified or congealed form" (Bernstein, 1971, p.46)

- An **orientation towards the future** (hope, intention, disposition), not the past (secure knowledge): "anticipation is ... more primary than recollection; projection than summoning of the past; the projective than the retrospective" (Dewey cited in Bernstein, 1971, 207).
- **An appreciation of (and contributions toward) consciousness as dynamic and emergent**, rather than static but wanting to collect more information. Consciousness is not a fixed form of mental seeing (i.e., disinterested understanding), but more along the lines of "the craftsman involved in doing and making.... The craftsman perfects his art not by comparing his product to some 'ideal' model, but by the cumulative results of experience—experience which benefits from tried and tested procedures, but always involves risk and novelty" (Bernstein, 1971, paraphrasing Dewey, p. 219)
- **Ongoing somaesthetic training, self-improvement, and self-stylization**. If experiences of art and beauty are distinctive for the powerfully gratifying ways they absorb our attention, unify our consciousness, and engage our emotions, then increasing our powers of awareness, focus, and feeling through better mastery of their somatic source could render more of our experience similarly rewarding in such ways. (Shusterman 21.1)

An aesthetic experience is one in which (a) the aesthetic goods listed above are *experienced* or *felt*, and also (b) the experience of them contributes to the long-term *somaesthetic skills* of insightful perception, imagination, meaningful self-stylization, and a disposition to do good.

All of this is not to suggest that every single interaction design must meet all of the above norms; rather, pragmatist holism would seem to suggest that interactions need only to contribute to and participate in lived ecologies where these qualities are experienced and subsequently cultivated through practice into skills. It is the lived world that ideally needs to have these qualities, not every single thing a person touches within the lived world. But inasmuch as our lived world is artificially designed—buildings, clothing, parks, appliances, furniture, and now interactive technologies—the burden is on us as designers to make that artificial world humane.

Somaesthetics offers normative criteria and a conceptual vocabulary to facilitate the design and evaluation of humane interactive products.

### 21.9.3 Criticisms and Limitations

It is probably obvious that I am sympathetic to somaesthetics and believe that it can contribute to HCI and interaction design. But one bad habit that we as a field have is that while we are eager to advocate for the introduction of a given theory, we often don't acknowledge that this theory has confounds or limitations. I have tried to argue some of the specific ways that somaesthetics can contribute to HCI, and I think it's also important to explore some of the ways it is not particularly well positioned to contribute to HCI. By exploring both strengths and limitations, as a field we can use theories more effectively and have some sense for when alternative theories are needed.

#### 21.9.3.1 Somaesthetics is only loosely coupled with methods

HCI is a field that likes its methods, and it's not clear how somaesthetics translates into methods. More fundamentally, it's not even clear whether somaesthetics should translate into methods, at least not in the sense that the term is used in the sciences. At stake is an epistemological disagreement about how best to produce knowledge. A traditional experimental methodology, such as that described by Tractinsky in his interaction-design.org Encyclopedia entry on Visual Aesthetics, isolates variables in controlled experiments—classic methods from experimental psychology. For Kutti at least, such an approach is the antithesis of how designers operate:

> "*Thus from the viewpoint of a designer, HCI people were not designers but "barbarians," uneducated technicians lacking any understanding of the aesthetics and complexity of the cultural filtration involved in a design. This suspicion was strengthened by the HCI people's obsession on methods instead of personal judgment.*"
>
> *-- Kuutti, 2009: p. 8*

Kuutti seems to establishing an exclusive opposition between expert judgment-based approaches and methods, a position also suggested by Greenberg & Buxton (cited earlier). Indeed, the very existence of methods seems to dumb scientists down into "barbarians." Kuutti's provocative language aside, it is easy to understand why scientists might view personal judgment as lacking any methodological rigor and thus barely any better than "mere opinion," and why designers might view scientific methods as replacing human judgment with mechanistic algorithm-like recipes, which would seem to be a form of intellectual "barbarism." Though it's easy enough to understand these caricatures, it's less clear whether they have any validity or in fact if they mainly just exaggerate differences. Dewey harmonized scientific and artistic approaches, but he did so by treating the sciences as an aesthetic practice, a position that might not appeal to practicing HCI scientists (though there is not as much daylight between Dewey's position and that of post-positivist science, e.g., Quine, as one might expect).

Regardless of how apparent or real the opposition is between expert judgment and scientific method, it's not clear to me that somaesthetics is going to resonate with interaction designers for whom data and methods are paramount.

#### 21.9.3.2 Somaesthetics says little about the content of actual experiences

In a recent paper, Hassenzahl, Diefenbach, & Göritz (2010), who operate within a cognitive approach to UX, criticize McCarthy & Wright's pragmatist take on UX as follows:

> "*With their emphasis on "values, needs, desires and goals," McCarthy and Wright (2004) are in line with accepted psychological theories (see Carver and Scheier, 1989), which understand action as being permanently shaped not only by the context and conditions on an operational level, but also driven by overarching, universal psychological needs. The question at hand, however, is what these "values, needs, desires" are. In fact, McCarthy and Wright (2004) seem to explicitly avoid any commentary on the content of "needs." This is due to a critical view of attempts to reduce, what they call "felt experience", to a set of generalized concepts.*"
>
> -- *Hassenzahl et al (2010): p. 354*

What Hassenzahl, Diefenbach, and Göritz are getting at is that there is an empirical dimension to experience, that is, users of a given design do have an experience *e*, it should be possible for research to discover the content of that *e*, and that McCarthy & Wright's approach categorically fails to address that question, because it offers few strategies to capture "what these 'values, needs, desires' are"—which is an empirical, rather than critical, question.

I think there is a valid point here. I do see value in McCarthy & Wright that Hassenzahl, Diefenbach, and Göritz apparently do not, but I can understand why they read McCarthy & Wright in that way.

And now I will also add that I think a somaesthetic approach to HCI basically has the same shortcoming for basically the same reasons. Again, while somaesthetics is strong at offering an account of how an individual trains or cultivates the self as a perceiver and expresser (not just a thinker), it offers fewer tools to try to understand the content of particular experience *x*, and yet designers do have reason to want to know that.

#### 21.9.3.3 Shusterman's stock of examples isn't particularly helpful

For an encyclopedia entry on interaction design, Shusterman not only used examples that take a lot of work for HCI readers to understand in the way that he wants them to, but he also missed opportunities to explore somaesthetic HCI with appropriate examples from the field. It is, of course, easy to explain this problem as a result of Shusterman's outside status. Nonetheless, it is a huge missed opportunity, not just rhetorically (in terms of his ability to persuade HCI readers to engage with somaesthetics) but also substantively (some examples from HCI surely would help everyone think more deeply about somaesthetic HCI).

An obvious starting point is the field of robotics, in particular robotic work for domestic settings and everyday contexts, where the robots are designed to be appropriately meaningful as computational bodies in everyday life. Robot designer Tatsuya Matsui, for example, "believes that robots are like flowers. They can be delicate and beautiful. They are endearing and need nurturing" (Hornyak, 2007), a somaesthetic concept that is undeniably visible in his work:





Figure 21.1 A-B: Robots designed by Tatsuya Matsui.

Beyond robotics, several areas of HCI have explored embodied computational artifacts in relation to human embodiment. Another work is “Soft-Spikey Mouse,” created by artist Youngsuk Altieri working with Heekyoung Jung and myself:

**IMAGE MISSING PLEASE PROVIDE**

Figure 21.2: “Soft-Spikey Mouse,” created by artist Youngsuk Altieri working with Heekyoung Jung and Jeffrey Bardzell.

So when Shusterman writes, “At this stage in somaesthetics research, we have only been concerned with somatic feelings of human bodies and thus with only one side of the HCI interaction. But, in principle, it may be possible to consider the somaesthetics of nonhuman somas, including computer bodies” it is clear that he simply has not yet engaged with the considerable amount of work in our field that has already been doing precisely that for decades.

As I imagine how such examples might influence Shusterman’s thinking, and how much I want to hear what he has to say about such work, it becomes clear to me that somaesthetics just might benefit as much from HCI as the other way around.

And I wouldn't have it any other way.

## 21.9.4 References


- Bannon, L., J. and Bødker, S. Beyond the interface: En-countering artifacts in use. In Carroll, J., M. (ed), *Designing Interaction*. Cambridge University Press, 1991, 227-250.
- Bardzell, J. (2012). Commentary on: Tractinsky, Noam (2012): Visual Aesthetics: in human-computer interaction and interaction design. In: Soegaard, Mads and Dam, Rikke Friis (eds.). "Encyclopedia of Human-Computer Interaction". Aarhus, Denmark: The Interaction-Design.org Foundation. Available online at http://www.interaction-design.org/encyclopedia/visual_aesthetics.html
- Bardzell, S., Bardzell, J., Forlizzi, J., Zimmerman, J., and Antanitis, J. (2012). Critical design and critical theory: The challenge of designing for provocation. *Proceedings of Designing Interactive System DIS'12*. New York: ACM.
- Bernstein, R. *Praxis & Action: Contemporary Philosophies of Human Activity*. U Penn Press, 1971.
- Cockton, G. 2008. Revisiting usability's three key principles. In *CHI '08 Extended Abstracts*. ACM, New York, NY, USA, 2473-2484.
- Dunne, A. (2006). *Hertzian Tales: Electronic Products, Aesthetic Experience, and Critical Design*. Cambridge, MA: MIT Press.
- Dunne, A., and Raby. F. (2001). *Design Noir: The Secret Life of Electronic Objects*. Birkhäuser, Basel, Switzerland.

- Greenberg, S. & Buxton, B. (2008). Usability considered harmful (some of the time). *Proceedings of CHI'2008*. New York: ACM.
- Guignon, C., & Hiley, D. (2003). *Richard Rorty*. Cambridge: Cambridge University Press.
- Hassenzahl, M., Diefenbach, & Göritz, A. Needs, affect, and interactive products — Facets of user experience. *Interacting with Computers 22* (2010) 353—362
- Hornyak, T. (2007). Matsui's graceful robots evoke human emotion. *The Japan Times Online*. Available online at http://www.japantimes.co.jp/text/nc20071205a1.html
- Jung, H., Altieri, Y., & Bardzell, J. (2010). Computational objects and expressive forms: a design exploration. In *Proceedings of the 28th of the international conference extended abstracts on Human factors in computing systems*. ACM, New York, NY, USA, 3433-3438.
- Kuutti, Kari. (2009). HCI and design: uncomfortable bedfellows? In Binder, Lowgren & Malmborg (eds.) *(Re)searching the Digital Bauhaus*. London: Springer, 43-59.
- Kuutti, Kari. (2001). Hunting for the lost user. *Cultural Usability*. Available online: http://www.mlab.uiah.fi/culturalusability/papers/Kuutti_paper.html
- McCarthy, J. and Wright, P. *Technology as Experience*. The MIT Press, MA, USA, 2004.
- Nelson H. & Stolterman, E. *The Design Way: Intentional Change in an Unpredictable World*. New Jersey: Ed-ucational Technology Publications, 2003.
- Shusterman, Richard. (2008). *Body Consciousness*. Cambridge: Cambridge University Press.

- Shusterman, Richard. (2000). *Pragmatist Aesthetics: Living Beauty, Rethinking Art*. Second edition. Oxford: Rowman & Littlefield.
- Shusterman, Richard (2011): Somaesthetics: Thinking Through the Body and Designing for Interactive Experience. In: Soegaard, Mads and Dam, Rikke Friis (eds.). “Encyclopedia of Human-Computer Interaction”. Aarhus, Denmark: The Interaction-Design.org Foundation. Available online at http://www.interaction-design.org/encyclopedia/somaesthetics.html
- Tractinsky, Noam (2012): Visual Aesthetics: in human-computer interaction and interaction design. In: Soegaard, Mads and Dam, Rikke Friis (eds.). “Encyclopedia of Human-Computer Interaction”. Aarhus, Denmark: The Interaction-Design.org Foundation. Available online at http://www.interaction-design.org/encyclopedia/visual_aesthetics.html
- Zimmerman, J., Forlizzi, J. and Evenson, S. Research through design as a method for interaction design research in HCI. In *Proc. of CHI’07* ACM Press (2007), 493-502.

## 21.10 COMMENTARY BY KRISTINA HÖÖK

How to cite this commentary in your report

### Kristina Höök

© *Kristina Höök*

Kristina Höök is a professor in Interaction Design at Royal Institute of Technology (KTH) in Sweden. She started and now works in the Mobile Life centre. She also upholds a part-time position at SICS (Swedish Institute of Computer Science)....

Kristina Höök
Kristina Höök is a member of The Interaction Design Foundation

In designing for bodily experiences, there has been a lack of theories that can provide the underpinnings we need to understand and deepen our design thinking. Despite all the work we have seen on designing for embodiment (Dourish, 2004, and others), the actual corporeal, pulsating, live, felt body has been notably absent from both theory and practical work. At the same time, digital products have become an integral part of the fabric of everyday life, the pleasures (and pains) they give, their contribution to our social identity, or their general aesthetics are now core features of their design. We see more and more attempts to design explicitly for bodily experiences with digital technology, but it is a notably challenging design task. With the advent of new technologies, such as biosensors worn on your body, interactive clothes, or wearable computers such as mobiles equipped with accelerometers, a whole space of possibilities for gesture-based, physical and body-based interaction is opened.

Some claim that the technologies we wear today treat our bodies in a negative way:

> "*Electronics, robotics, and spintronics invade and transform the body and, as a consequence of this, the body becomes an object and loses its remaining personal characteristics, those characteristics that might make us consider it as the sacred guardian of our identity.*"
>
> *-- Longo, 2003*

How can we do a better job in interaction design involving our bodies — *the sacred guardians of our identity*? This is where I think Shusterman's theories of somaesthetics are relevant.

### 21.10.1 Three questions: What experiences? Articulation? Experiential qualities?

To design for corporeal, bodily, movement-based interactions, speaking to our senses and aesthetic experiences is difficult. Three questions immediately pops to my mind. First, what kinds of subjective, pleasurable or displeasurable, experiences are we aiming to design for? Glossing them over as all being about designing for *flow* (Csikszentmihalyi, 1990) or good *gameplay* experience is too vague (as we argue in Isbister and Höök, 2009). We need to drill deeper and better understand exactly what experiences we are talking about. Are we designing for pleasurable or unpleasurable ones? Are we designing for those that are subjective and unique, or ones that are common and shared? Ones that deliver serendipitous experiences or ones that are evocative and emotional? These are not all the same, even if they all emphasise aspects of bodily experience.

> "*A particularly difficult issue lies in understanding how these kinds of experiences may unfold over time — both in the particular interaction with and manipulation of the artefact but also as parts of our everyday on-going lives. As Löwgren puts it, a gestalt for interactive artefacts is defined as a "dynamic gestalt" which "we have to experience as a dynamic process.*"
>
> -- *Löwgren, 2001: p. 35 - 36*

Second, once we know what kind of experience we are aiming to design for, we need to *articulate* them in a form that makes sense and that we can share within a design team. Ways of knowing can arise from your bodily acts without any language translation in-between. The feel of the muscle tensions, the touch of the skin, the tonicities of the body, balance, posture, rhythm of movement, the symbiotic relationship to objects in our environment — these come together into a unique holistic experience. It is not the ability to fulfil a task, but the experience of the corporeality of doing so that matters here. Those descriptions also need to be shared with the users that we invite to test our designs, or even participate in the design process.

Thirdly, if we try to design interaction that builds on bodily movement, seeking certain experiential qualities, many different aspects of the interaction have to be fine-tuned to enable the experience, as, for example:

- the *timing* of interaction: movement has to render response in exactly the right moment for exactly the right kind of length of time in order to create for a particular experience (Sundström et al., 2005)
- linking *emotion and movement*: certain movements and body postures are more likely to coincide with certain emotional experiences (Darwin, 1872, Sheets-Johnstone, 1999, Laban and Lawrence, 1974, Moen, 2006, Paiva et al., 2003)
- *harmony of modalities*: the modalities of the interaction, such as graphics, haptics or gestures, all have to speak together — harmonize (Ståhl, 2006)

Before turning to what Shusterman and his work on somaesthetics can provide us with here, let me remind just give a few examples of how bodily interactions have been seen in the field of HCI or interaction design.

## 21.10.2 Strands of interaction design dealing with bodily interactions

### 21.10.2.1 Ergonomics

In ergonomics (preceding HCI — see Grudin 1990), the actual physical body is the core focus. The body has been measured and designed for in spaces such as airplane cockpits, cars or nuclear plant control rooms. As pointed out by Harper and colleagues, the perspective taken is one where humans are seen as part of a machine. The pilots, car drivers and factory workers are part of a larger machinery. They must be trained to follow certain routines automatically as if they are one part of the machine. The machinery must be fine-tuned so that human error is minimized and this can only be done through designing the machinery to fit with meticulous measurements of our physical capacity. In those situations, we want to see our bodies as machines, able to follow routines and act in error-free ways in the spur of a moment (Harper et al., 2007).

While it may sound negative to take such a narrow view on the body, treating it as a machine, we must remember that sometimes we really do want to see ourselves as machines. It is of key importance to us that risks are minimized with driving a car or controlling a nuclear plant. Ergonomics has also cared for the body, aiming to avoid harming the body. By changing the way a machine works, its users can fit better into the machinery. But in ergonomics, for the most part, we assume the body to be passive — the interface will be sending signals to the human body that the passive body receives, sending onwards to the way more important mind. The body is not a subject, actively perceiving and acting.

#### 21.10.2.2 Cyborgs

Another position towards the body sometime taken in HCI is that of *cyborgs*. A cyborg consists of both artificial and natural systems, or to phrase it differently, of both human body and designed tools that extends out capacity. In its simplest form the extension can be the stick that a blind man uses to find his way. The stick becomes a part of how he feels the world, an *embodied* part of his own body. But framing tools as part of our cyborg existence goes beyond this one- way extension of our bodies. The cyborg concept comes with various ethical and moral implications when we regard how the technical tools we extend our bodies with in turn speak back to us. This positive side of being a cyborg is in some sense that we can free ourselves from our bodies — as discussed by the feminist Donna Haraway in her cyborg feminist writings (1991). The focus in this movement is on extending the mind, freeing us from our corporeal reality.

While this body-less cyborg being on the internet was much discussed in the beginning of the virtual reality-era, the pendulum has now swung back and most regard it as bad behavior to not connect your real identity to your virtual identity. In addition, more and more technologies are tying reality and virtuality more strongly together, entering our physical selves into the virtual spaces. For example, in the computer games area, we have new interaction devices, such as WII, fake guitars in Guitar Hero, or mobiles, connecting more strongly with our physical selves.

#### 21.10.2.3 Trimming the body

A growing body of work, focuses on the body itself as the domain or the focus of attention. Here, HCI focuses on interactions for sports, healthy living, or physical activity. These systems often treat our bodies as objects that we can study from the outside, that can be trimmed and controlled. Again, the body becomes subordinate to mind, as an instrument or machine, passively receiving sign and signals, but not actively being part of producing them.

These kinds of systems may have many benefits; relieving our bodies from pain, creating interesting experiences, or making us healthy. At the same time, by making the body into a machine that can be measured and studied as an object, we risk putting 'goals' and 'tasks' to our bodies without turning to our felt life. Pedometers measure how many steps we take and the goal becomes to walk at least 10.000 steps per day — not matter how we feel about walking that particular day. Again, many of these systems reinforce a dualistic stance where the body is a separate entity that can be measured and dealt with as an object. It is not the sensory-locus of ourselves.

#### 21.10.2.4 Third wave

In the "third wave" of HCI, design for experiences goes beyond those of task completion, efficiency, and tool-based perspectives. This includes designing for bodily experiences. So far, when it comes to involving bodies and creating for bodily experiences, the focus has mainly been on sports and games (e.g. from early work (Ishii et al., 1999) to current (Benford et al., 2012)). The aim is to design for experiential qualities such as flow, immersion or uncomfortable experiences. But there is also a growing body of designs aimed at other experiences. One example is Moen's Body Bug — a wire that you wrap around your body where a 'bug' registers your actions and climbs up and down a wire (Moen, 2006). The bug is a simple robot, moving along the wire. When you strap the wire around your body and start making movements, the bug will move along the wire, in a sense mirroring your movements. The bug makes you want to 'dance'. The sought experiential quality is that of enjoying your own body movement as we do when we dance.

To reach designs in which such qualities arise, designers and researchers have repeatedly reported that as designers, we need to experience our own bodies in the design process (Hummels et al., 2007). This in turn requires new methods in the design process.

Recently, we have started to see other studies where HCI researchers attempt to observe different cultures or communities of practice, for insight on how to design for novel bodily experiences. There are ethnographic studies about hunting culture (Juhlin and Weilenmann, 2008), skaters and golfers (Tholander and Johansson, 2010), horseback riding (Höök, 2010) to citizens constrained by electronic surveillance bracelets (Troshynski et al., 2008). These studies repeatedly tell us that bodily experiences have been undervalued in ICT design and that there is little knowledge on how to address them.

The study by Tholander and Johansson (2010), on skaters and golfers show that those practices do not distract their users from being in the world together with their skateboards or golf clubs. Tholander and Johansson convincingly argue that interactive technologies that aim for physical interaction too often force users to interact through some type of screen interface, taking away focus from the environment. Instead of interacting with others around us or with the surrounding nature, we focus on the screen.

The study by myself on horseback riding, (2010) I try to provide a rich account of how horseback riding involves all our senses, at moments involving us in *centaur*-experiences — feeling as one with horse and environment. My point is to show both how impoverished interaction with many of our interfaces are compared to the sensory richness of riding, and also how impoverished our articulations of interactions are, the lack of an agreed upon language for describing interactions.

Troshynksi and colleagues, in their study on paroled sex of- fenders who are required to wear a GPS tracking electronic bracelet on their ankle (2008), show how this technology constrains their bodies in ways beyond that of the original intent of the technology. A considerable amount of work is put into preserving the technology intact during everyday routines like showering, and their mobility in the environment is considerable constrained, among other implications.

All of these studies point to limitations in the ways we think of today's wearable and mobile technologies and their impact on bodily behaviors and practices.

### 21.10.3 What Shusterman brings to the table

From this brief walk through some of the work involving our bodies in digital interactions, we can now turn to Shusterman's work and perhaps see a bit more clearly why his theories on somaesthetics are appealing to some of the interaction design researchers in our field.

When designing for a non-dualistic stance towards body and mind, we need some way of talk about what experiences we strive to engage ourselves and our users into. While most accounts of corporeal involvement will be mainly descriptive, Shusterman's somaesthetics is also *normative*. He tells us that by engaging in certain practices, in inward listening and learning, we can know ourselves better, and thereby understand and interact with others more fully. It trains our empathy — both with ourselves and others. While this may all sound mysterious and fluffy, the take away message is, in my view, not religious or mysterious. It simply says that we can train our bodies, our muscles, our nervous system (including the brain), to become more knowledgeable and aware of ourselves. As I am a horseback rider, I know that any predominately movement- and body-based practice requires this kind of training and knowledge. As mentioned above, I have tried to describe the complexity of knowledge required to ride a horse in an autoethnographic study (Höök, 2010). The interaction with a horse is obviously not word-based. It happens through physical signs and signals: the riders use the muscles in their legs, the placement of their sitting bones, bodily balance, head movement, hand and arm connection to the horse's mouth and sometimes tone of voice. The horse talks back through its movement, direction, pace, activations of muscles that can be felt throughout the horse's body, its head movements, tail movements, flipping ears, bend of neck and noises. In order to be a good rider, you need to learn this wordless language. As in any language, understanding and communication arises in interaction over time. When you have experienced a particular bodily schema or concept yourself this understanding may arise.

When designing digital interactions, we should be able to articulate, shape and design for equally detailed descriptions of movement, body and physical signs and signals flowing back and forth between us and the system we are designing. More importantly, as Shusterman points out, moving your body is not only a matter of performing a function, it is also an *aesthetic* experience. There is a plenitude of activities that we do for the pleasure of moving — dancing, sports, jogging, cycling. The pleasures of these activities are of course not only soft, flowing movements, since some of the activities involve pain, applying yourself really carefully to make your body do them, adjusting your own body in various ways, even making your body build certain muscles that you normally do not use so much, embarrassment when you do not get it right, and so on.

Interaction design has perhaps been a bit too obsessed with zero-learning time, an issue that will not sit well with some of the movement-based practices Shusterman is advocating. A take-away message from Shusterman is that it takes time to learn. You have to apply yourself. Getting to know yourself, your own body, changing your movements, training yourself, is not “natural” — even if your body is “there for you” all the time. Similar to how you must learn to think and reason, you must learn how to listen to your body, how to improve your body knowledge.

### 21.10.4 Turning to design

As pointed out by Bardzell in his comment to this chapter, the translation from theories of somaesthetics into HCI and interaction design is non-trivial. Obviously, any body will have different parts (legs, arms, brain, nervous system) and different processes in that body will have different characteristics, but they are intimately linked. Likewise, bodies move through difference spaces, social and physical, shaping interactions. Or as put by Suchman picking up on Latour, unavoidably a part of complex temporal, material and social assemblies, the body is unceasingly (re)configured in relational terms. We perceive, act and understand the world as unities of mind, body, routine, culture, social settings and with machines as part of our ways of being in the world. The design process needs to consider the connections between these processes, moving beyond a narrow focus on cognition as it happens in our brains.

In particular, I have been interested in the linking from movement to emotion and back. Early on Darwin made a strong coupling between emotion and bodily movement (Darwin 1872). Since then, researchers in areas as diverse as neurology (LeDoux 1996; Davidson et al. 2003) to philosophy and dance (Laban & Lawrence 1974; Sheets-Johnstone 1999) describe the close coupling between readiness for action, muscular activity and the co-occurrence of emotion. Sheets-Johnstone makes the case that:

*"Without the readiness to act in a certain way, without certain corporeal tonicities, a certain feeling would not, and indeed, could not be felt, and a certain action would not, and indeed, could not be taken, since the postural dynamic of the body are what make the feeling and the action possible.( "*

*-- Sheets-Johnstone 1999: p. 265*

Or as Dewey puts it:

"There is, therefore, no such thing in perception as seeing or hearing plus emotion. The perceived object or scene is emotionally pervaded throughout."

If we attempt to define a lived emotional experience in dualistic terms, we will surely fail, but with a perspective where man is seen as a whole, both body and mind, both individual and part of the world, the gulf between our interpretative experiences and what can/cannot be studied will not be as problematic.

Just to make it slightly clearer, here is an example of such a co-occurrence of emotion and movement from my horseback riding account:

"As horse and rider move together, they create a rhythm. Depending on the gait, it can be a two-beat (trot, pace), three-beat (canter), or four-beat (walk, gallop, tölt), in different paces. To allow the horse to keep the beat in a balanced way, the rider needs to make herself invisible in the saddle, not disturbing the rhythm. [..] The problem was that I was sitting back into the saddle with a 'splat' slightly out of rhythm with the horse. Given how many years I had been riding before going to lessons with Christian [my horseback riding teacher], it was horribly embarrassing for me to be out of synch. Following the rhythm of the horse is one of the most important pleasures of riding. As discussed by others (Moen, 2006), rhythmic movement as in dance or riding, moves us in way which are immediately appealing. But just as it can be very awkward to watch someone dancing out of rhythm, it is very awkward to experience it. [..] The embarrassment came from the actual physical experience of being out of rhythm. Our bodies are used to rhythms, our own bipedal swagger (Sheets-Johnstone, 1999), our mother's heart beat, waves beating the beach, music and dancing, and, for those who are fortunate enough to experience it: the horses' different gaits."

Translating from this experience in horseback riding, we can see many possible digital interactions picking up on rhythm. Take for example the work by Danielle Wilde named hipDisks recently exhibited at CHI 2012 (Wilde, 2012):

> "*Possibly the most undignified musical instrument ever, hipDisk exploits changing relationships between torso and hip to actuate sound. Simple horizontal disk-shaped extensions of the body exaggerate, so make highly visible, the interdependent relationship of the hip and torso. Soft switches, strategically placed around the perimeter of each disk, allow the wearer to play a chromatic scale, and so play simple melodies, restricted only by flexibility and speed of swing* "
>
> -- *Wilde, 2012*



FIGURE 21.1 A-B-C: Thecla Schiphorst dancing with Danielle Wildes' Hipdisk.

A designer that has picked up more directly on somaesthetics (and who is also commenting on this chapter) is Thecla Schiphorst. She suggests interactions and design methods that require particular movements, such as moving very, very slowly in order to listen to your own bodily state in interaction or attaching users by velcro asking them to move and interact together in order to explore extensions of the body and their meaning in terms of privacy (Schiphorst, 2007). She has also built a couple of systems, like soft(n), that explores the somaesthetics of touch and interaction through interactive artifacts:

> "*The soft(n) installation is an intelligent tangible network comprised of soft physical objects that exhibit emergent behavior through interaction. soft(n) is a group of 10 interactive soft objects, each containing a specially designed and custom-engineered multi-touch soft input surface and motion detectors. Each soft object has an ability to actuate vibration, light and sound in response to its tactile induced state.*"
>
> -- *Schiphorst, 2009*

Another compelling example is the work by Høbye and Löwgren on the system named Mediated Body (2011). A performer and a participant both wear earphones and through turning them into human antennas, they can generate evocative music when they touch each-others' bare-skin or "auras". Again, this invites a somaesthetic, in this case, social experience.

Figure 21.2: Richard Shusterman and Kia H??k using the 'Mediated Body' system at the CHI 2012 conference.



In my view, apart from using Shusterman's somaesthetic theories to train ourselves as designers or train our users to express what they experience (as discussed by Bardzell in his commentary), I believe that we can build some of the ideas into the actual designs of interactive systems.

### 21.10.5 In summary

In a sense, the interest in emotional experiences and third-wave HCI has served as a bridge for the whole field of HCI to turn from symbolic, analytical ways of doing task analysis and designing for efficient ways of supporting tasks, to caring more about experiences in general. It has also, to some researchers in HCI, served as a bridge to start addressing our physical, corporeal bodies in interaction and to attempt to bridge the dualism chasm.

This has, in turn, created a huge space of opportunities for design that puts our bodily ways of being in the world first and attempt to address our corporeal experiences. The systems we have been designing in my group (eMoto, Affective Health and others — turn to the chapter on Affective Interaction in the Encyclopedia of interaction-design.org for a longer description) have all been attempts to address the interaction between emotion and movement. While each of these systems has its deficiencies, none of them is trying to reduce human experience to something that can be measured and modeled, and then packaged as an information piece to be sent to others. They are “non-reductionist” (Höök et al., 2008). The experience of using them emotionally and corporeally is shaped by the participants. In a sense this becomes the “participatory design”-movement of the third wave of HCI (Höök, 2006).

It remains to be seen how we can translate the insights from Shusterman’s work on somaesthetics into design. And perhaps, we will have to look for other, complementing theories of bodily interactions, sometimes with less normative perspectives on what is good and what is bad, and perhaps with a stronger orientation towards our socio-bodily practices. Most of all, if we continue to create interactions that come closer and closer to our bodies, our “sacred selves”, we need to be guided by some values or ideas of why and how to do so. The world is flooded with crappy technologies that harm our bodies and our means to be corporeally present in the world, together with others. Addressing aesthetics of the soma, also means addressing important values in design.

### 21.10.6 References


- Benford, S., Greenhalgh, C., Giannachi, C., Walker, B., Marshall, J., and Rodden, T. (2012) Uncomfortable Interactions, in Proceedings of *30th CHI Conference on Human Factors in Computing Systems*, Austin, Texas, ACM Press.

- Csikszentmihalyi, M. (1990) *Flow: The Psychology of Optimal Experience*. Harper Perennial, London.
- Darwin, C. (1872/1998) *The expression of emotions in man and animals*, Third ed. by Paul Ekman. Oxford University Press.
- Dourish, P. 2004 *Where the action is: the foundations of embodied interaction*. Cambridge, MA: MIT Press.
- Grudin, J. (1990). The computer reaches out: The historical continuity of interface design. Proc. CHI '90, 261-268.
- Haraway, D. (1991) A Cyborg Manifesto: Science, Technology, and Socialist-Feminism in the Late Twentieth Century, in *Simians, Cyborgs and Women: The Reinvention of Nature* (New York; Routledge, 1991), pp.149-181.
- Harper, R. Randall, D. Smyth, N. Evans, C. Heledd, L. And Moore, R. (2007). Thanks for the memory, In *Proceedings of Interact: HCI 2007*, Lancaster, Sept.
- Hummels, C., Overbeeke, K. C., and Klooster, S. (2007) Move to get moved: a search for methods, tools and knowledge to design for expressive and rich movement-based interaction, In *Personal and Ubiquitous Computing*, Volume 11, Issue 8, Pages: 677 — 690, Springer-Verlag London, UK
- Høbye, M., & Löwgren, J. (2011). Touching a stranger: Designing for engaging experience in embodied interaction. *International Journal of Design*, 5(3), 31-48.
- Höök, K. (2006). Designing Familiar Open Surfaces, In *Proceedings of NordiCHI 2006* Oslo, Norway, October 2006, pp. 242 — 251, ACM press.
- Ho?o?k, K. (2010) Transferring Qualities from Horseback Riding to Design. *Proceeding of the 2010 Nordic conference on Human factors in computing systems*, *NordiCHI 2010*, Reykjavik, Iceland.

- Höök, K., Ståhl, A., Sundström, P., Laaksolahti, J. (2008) Interactional Empowerment, In Proceedings of *ACM SIGCHI conference Computer-Human Interaction* (CHI2008), Florence, Italy, ACM Press.
- Ishii , H., Wisneski , C., Orbanes , J., Chun, B., Paradiso, J. (1999) PingPongPlus: design of an athletic-tangible interface for computer-supported cooperative play, *Proceedings of the SIGCHI conference on Human factors in computing systems*, p. 394-401, ACM Press.
- Juhlin, O. and Weilenmann, A. Hunting for fun (2008) In *Proceedings of the 2008 ACM conference on Computer supported cooperative work*, pp. 57-66.
- Laban, R. and Lawrence, F.C., (1974). *Effort Economy of Human Effort*, Second ed. Macdonald & Evans Ltd., London, UK.
- Longo, G. O. (2003) Body and Technology: Continuity or Discontinuity? In Fortunati, L., Katz, J. E., & Riccini, R. (eds.) *Mediating the human body: Technology, communication and fashion*. Mahwah, NJ : Lawrence Erlbaum & Associates.
- Löwgren J. (2001). From HCI to Interaction Design. In Chen, Qiyang (ed.), *Human-Computer Interaction: Issues and Challenges*, Hershey, PA, USA: Idea Group Inc., 2001.
- Moen, J., (2006). *KinAesthetic movement interaction: designing for the pleasure of motion*. Unpublished Doctoral Thesis, KTH, Numerical Analysis and Computer Science, Stockholm, Sweden.
- Paiva, A., Costa, M., Chaves, R., Piedade, M., Mourão, D., Sobral, D., Höök, K., Andersson, G., and Bullock, A. (2003) SenToy: an Affective Sympathetic Interface, *International Journal of Human Computer Studies*, 59:1-2, pp. 227-235, Elsevier.

- Schiphorst, T. (2007). Really, really small: the palpability of the invisible, C&C '07: Proceedings of the *6th ACM SIGCHI conference on Creativity & cognition*, ACM Press.
- Schiphorst, T. (2009) soft(n): toward a somaesthetics of touch. CHI '09: *Proceedings of the 27th international conference extended abstracts on Human factors in computing systems,* ACM Press, pp. 2427—2438.
- Sheets-Johnstone, M. (1999) Emotion and Movement: A beginning Empirical-Phenomenological Analysis of Their Relationship, *Journal of Consciousness Studies*, 6, No. 11-12, pp. 259-277.
- Ståhl, A., (2006). *Designing for emotional expressivity*. Unpublished Licentiate Thesis in Design, Umeå Design Institute. Umeå University, Sweden.
- Sundström, P., Ståhl, A., and Höök, K. (2005) eMoto - Affectively Involving both Body and Mind, In *Extended abstract CHI2005*, Portland, Oregon, USA.
- Tholander, J., and Johansson, C. (2010) Design qualities for whole body interaction: learning from golf, skateboarding and BodyBugging. In *Proceeding of the 2010 Nordic conference on Human factors in computing systems, NordiCHI 2010*, Reykjavik, Iceland, pp. 493-502.
- Troshynski, E., Lee, C., and Dourish, P. (2008) Accountabilities of presence: reframing location-based systems. *Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, pp. 487-496.
- Wilde, D. (2012) hipDisk: Experiencing the Value of Ungainly, Embodied, Performative, Fun, Interactivity exhibition at CHI 2012, Austin, Texas, ACM Press.

## 21.11 COMMENTARY BY ERIK A. STOLTERMAN

How to cite this commentary in your report

### Erik A. Stolterman



Erik Stolterman is Professor of Informatics and Department Chair in the School of Informatics and Computing, Indiana University, Bloomington. Stolterman's main work is within interaction design, philosophy and theory of design, information technology and society, information systems design, and philosophy of technology. Stolterman has published a large number of articles and five books, ...

Erik A. Stolterman
Erik A. Stolterman is a member of The Interaction Design Foundation

Humans have bodies. Bodies are not only the physical mechanical carriers of who we "really" are. Humans interact with and through their bodies. Users are humans and have bodies. Human computer interaction is therefore about bodies as well as cognitive and emotional minds. All this has become a concrete practical reality over the last decade to anyone involved in interaction design. New interactive technology has changed interaction away from being purely representational (commands and text) to direct and bodily interactions (touch, haptic and gesture) and away from being purely screen based to being embedded in and parts of our physical artifacts and environments.

This development means that interaction designers are more than ever before challenged by bodily aspects of interaction. This has of course been addressed both empirically and theoretically within our field through notions such as, ubiquitous computing, embodied interaction, tangible interaction, haptics, gestures, etc.

In contrast to what is common in our field, Shusterman delivers a more profound approach to the challenge of bodily engagements with technology. As a philosopher and as the founder of the term somaesthetics, Shusterman has over the years, in a serious and foundational way, developed an "integrative conceptual framework" for a better understanding of "somatic experiences" (quotes from his article).

What Shusterman can do, that few in our field can do, is to ground the phenomena and ideas related to bodily interactions in a historical and philosophical context and scholarship. This is extremely helpful for our field and it pushes HCI research forward to a more developed understanding.

It is hard to argue against the basic position that Shusterman present. He writes: "By exploring the fundamental features of our embodied ways of engaging the world and transforming it through action and construction, somaesthetics can provide useful insights and experiential skills to help designers produce products and situations that provide more rewarding and pleasurable experience." Few would argue against the idea that we can obtain other insights about our reality when we "use" our body as a sensory "tool" than if we solely approach it through intellectual and cognitive means.

Reading Shusterman's article is interesting and potentially useful for anyone involved in interaction design, especially when engaged with any kind of physical aspects of interactive artifacts. However, while Shusterman delivers an excellent account of the theoretical and philosophical aspects of embodiment and somaesthetics, I am somewhat disappointed when he more directly tries to describe how this can be applied or used in interaction design.

Shusterman does this by bringing somaesthetics aspects not only to the relationship between a user and an artifact but into the design process and to the thinking and doing of a designer. He writes for instance "... the body is our indispensable tool of tools" in a design process. He continues "... designers could improve their design skills by becoming more aware of how they use themselves and how they feel when using particular products rather than merely thinking of how the product is conventionally used." Of course, it can be argued that this is already common practice in interaction design, since prototyping already is a core activity in our field. Working with material prototypes at any level of fidelity means that the designer or user engages in bodily explorations of ideas. Prototypes are in most cases physical manifestations of design concepts that are developed explicitly with the purpose to explore precisely what Shusterman suggests, namely how a designer or user "feel when using a particular product rather than merely thinking of how the product is conventionally used".

It might be fair to acknowledge that when Shusterman talks about using the body as a tool he is in many ways more bodily oriented than what is the case when exploring prototypes. For instance, he also sees bodily engagements as a way to explore more abstract ideas concerning potential design directions. However, the idea of using the body as a tool in design activities has also been more directly explored in our field through techniques such as "body storming" and similar methods. So, while Shusterman provide a philosophical foundation for many of these bodily oriented design process activities, it is also clear that designers are already engaged in somaesthetic activities as a way to better understand design ideas and user experience.

Shusterman mentions a few examples of what designers can do to develop a more somaesthetic understanding of their prototypes and designs. But again, for many interaction designers these examples are quite similar to what they already do during prototyping, design sessions, evaluations, etc. For instance, Shusterman writes that "A comparative study of how different shapes of cup handles affect one's feelings towards drinking, for example, could be used to improve cup design. Does the handle make your forefinger grip tightly, and how does this affect the rest of your arm, the rest of your body, and by extension, the feeling of drinking?" This example could be found in any interaction design textbook on the notion and evaluation of user experience design. What Shusterman proposes is if phrased in the everyday language of interaction design practice a matter of prototyping and contextual evaluation of artifacts in use. The basic belief in interaction design is that since these artifacts are objects and have physical properties, their properties influence the whole user experience and to develop an understanding of that experience, of course the artifact has to be examined in a way that includes the embodied aspects of interaction. So, with this example Shusterman is not really opening up anything new to most interaction designers, instead the example can probably backfire and make the theoretical contribution to appear less interesting.

Another example that Shusterman mentions is HRI (human robot interaction). This is an area where a lot of research is already focused on how people experience the way robots move and in particular how their "bodies" are designed and how humans can, want to, or refuse to (bodily or otherwise) interact with these artificial bodies. For instance, a lot of research has been devoted to the bodily aspect of facial expressions, arm movements, even the embodiment of free roaming "smart" robot vacuum cleaners. In HRI the somatic aspects of design becomes almost unavoidable. There is no traditional interface, there is very little interaction related to disembodied intellectual and cognitive aspects through representational interaction. Instead, robots are themselves embodied and the interaction with humans is embodied from both sides. So, it can be argued that HRI is not, as Shusterman suggests, a potential area for somaesthtics, instead it is a field where somaesthetic approaches are unavoidable and already in practice, even if not theoretically refined.

The article of Shusterman exemplifies something that is quite common when it comes to design (research) and that is the difficulty of transforming advanced theoretical constructs into relevant and practical support for actual design work. We have over the years seen several examples of theoretical and philosophical approaches that have been both recognized and influential in academia as important scholarly contributions. However, many of these, such as Activity Theory, Distributed Cognition, and others, have proven to be difficult to translate into recommendations for practice and few have successfully reached a broader audience among practitioners. Yvonne Rogers (2004) offers an excellent account of this situation, with numerous examples and explanations of why this is a challenge to the field. Another treatise of the same topic can be found in my article (Stolterman, 2008) where I introduce the notion of "rationality resonance". This concept manifests the idea that any theory, to be practically relevant, has to be based on a deep understanding of existing practice. There has to be a resonance between the existing rationality in practice and the "new" rationality manifested in the proposed theory. This is of course not an issue for theories that only claim to contribute to our understanding of the field, but as soon as a claim is made that a theoretical construct is "useful" in practice this becomes a critical issue. The rationality embedded in a theory with such a claim must resonate with existing rationality in practice, that is, the theory needs to be based on a deep understanding of practice in all its richness and complexity. This means that the proposed rationality has to resonate with or at least pay respect to every aspect relevant in practice, such as, management, resources, time, skill and competence, even if these aspects are not core to the theory.

The theoretical and philosophical foundation that Shusterman offers is, in my view, excellent and should be required readings for any designer engaged with embodied interaction. But I find that the proposed theory or more precisely the suggested practical applications does not yet show enough resonance with existing practice or pay enough insight and respect to the complexity of existing practice. At the same time, it is unclear to what extent the practical design side of somaesthetics isn't already practiced. Maybe what many interaction designers are already doing in their serious attempts to capture people's overall experience of interacting with artifacts are already examples of a somaesthetic approach. If so, then Shusterman's contribution is not to be evaluated in relation to how "useful" it is but to how well it establishes a scholarly and philosophical foundation that existing practice may relate to and rest upon. If that is the case then instead of reading somaesthetics as an approach for design it could be used as a suitable tool for analyzing and understanding existing practice. But, if somaesthetics is actually meant to be seen as something radically different when transformed into practice then we are still looking forward to that to be developed and explained.

### 21.11.1 References


- Rogers, Y. (2004) New Theoretical Approaches for Human-Computer Interaction, Annual Review of Information, Science and Technology, vol. 38 pp. 87-143.
- Stolterman, E. (2008). The nature of design practice and implications for interaction design research. International Journal of Design, 2,1 55-65.

## 21.12 COMMENTARY BY YOUN-KYUNG LIM

How to cite this commentary in your report

### Youn-kyung Lim



Youn-kyung Lim is Associate Professor at the Department of Industrial Design at KAIST, South Korea. She is currently in CMU HCII in US as a visiting scholar for her sabbatical year. Her research interests include experience-centered design, creativity and design, prototyping in design, interaction design theory, design informatics, and CSCW....

Youn-kyung Lim
Youn-kyung Lim is a member of The Interaction Design Foundation

It was around 2006 when I first encountered the term, somaesthetics. As a part of my research activities, I was searching for the most appropriate philosophical discourse for human dynamic bodily experience. At that time, I was developing a new approach for understanding and designing the invisible but tangibly experiential qualities of interactive artifacts. When I came across the concept of somaesthetics, developed by the Richard Shusterman, I became fully convinced that the idea of designing "interactivity qualities" is valuable and furthermore, that somaesthetics may form the core for further developing this idea, more specifically through interactivity attributes (Lim, et al., 2007; Lim, et al., 2009; Lim, et al., 2011).

For a HCI design researcher like myself – who very much appreciates the perspective of aesthetics of interaction – the most intriguing part of the concept of somaesthetics is that it provides the conceptual framework for *consciously* explaining the "sense of quality" that is sourced and experienced from our bodily senses when we interact with interactive artifacts. Somaesthetics is not merely about bodily experience but more about the *articulation* of such experience. It is about *body consciousness* (Shusterman, 2008; Shusterman, 2011). The excerpt from one of my papers (Lim et al. 2011) below may describe the primary effect of such consciousness in the perspective of quality-centered design.

> "*The body becomes a tool for discovering new experiential spaces-but making the body conscious of what it experiences, and able to articulate that consciousness, is critical. Through the help of increasing consciousness, the quality sensibility can be increased, and this will lead to the experience of a higher aesthetic quality*"
>
> -- *Lim et al. 2011: p.115*

Somaesthetics is often applied in theatrical contexts to analyze performers' somatic styles of movement and postures, as Richard Shusterman mentions. Shusterman also notes that somaesthetics is not only about the performers but also about the observers of the performers. As a thought experiment, let us substitute the performers with interactive artifacts and the observers with users, or vice versa. Such mapping provides a new way of thinking about the relationship between a user and an interactive artifact. More specifically, such mapping allows us to focus on the *somatic* styles of movements and behaviors of the interactive artifact while interacted with by the user. Conversely, the interactive artifact may also interpret the user's movements and behaviors based on his or her somatic styles and respond to these in an appropriate way. As interaction designers, we can become conscious about this somatic relationship in our designs, and thus open a new space to be explored. It may also be appropriate in the context of robotics design as well.

In this way, the philosophical concept of somaesthetics extends the discourses around - and understandings of - our current concepts in Human-Computer Interaction. Somaesthetics may move our concepts in new directions. To me, somaesthetics is a very strong and effective philosophical concept to be learned and applied in interaction design especially when we are interested in *experienced* qualities of interaction. When we face emerging interaction technologies such as touch interfaces and gestural interfaces, this philosophical concept may, in my opinion, assume an even stronger role in Human-Computer Interaction.

### 21.12.1 References


- Lim, Y., Stolterman, E., Jung, H., & Donaldson, J. (2007). Interaction gestalt and the design of aesthetic interactions. *Proceedings of the 3rd Conference on Designing Pleasurable Products and Interfaces* (pp. 239-254). New York: ACM Press.
- Lim, Y., Lee, S., & Lee, K. (2009). Interactivity attributes: A new way of thinking and describing interactivity. *Proceedings of the 27th International Conference on Human Factors in Computing Systems* (pp. 105-108). New York: ACM Press.

- Lim, Y., Lee, S., & Kim, D. (2011). Interactivity attributes for expression-oriented interaction design. *International Journal of Design*, 5(3), 113-128.
- Shusterman, R. (2008). *Body consciousness: A philosophy of mindfulness and somaesthetics*. Cambridge, UK: Cambridge University Press.
- Shusterman, R. (2011). Somatic Style. *The Journal of Aesthetics and Art Criticism*, 69 (3), 147–159.


## 21.8 BEHIND THE SCENES

*Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.*

Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.

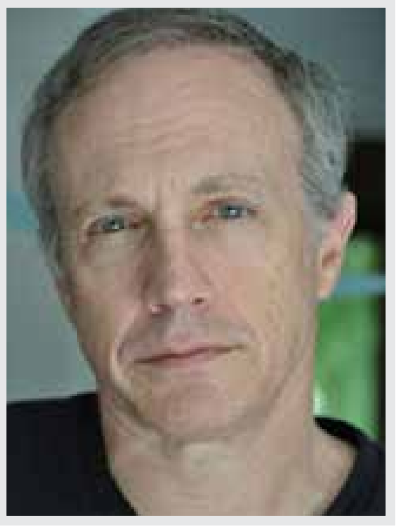
Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.

## 21.13 REFERENCES

Shusterman, Richard (2012): *Thinking through the Body: Essays in Somaesthetics*. Cambridge University Press

### About the author

**Richard Shusterman**



*© Richard Shusterman*

Current place of employment:
Florida Atlantic University

Richard Shusterman is an American pragmatist philosopher, currently the Dorothy F. Schmidt Eminent Scholar in the Humanities and Professor of Philosophy at Florida Atlantic University, where he directs the Center for Body, Mind, and Culture. He is internationally known for his contributions to philosophical aesthetics and pragmatism.

## YOUR NOTES AND THOUGHTS ON CHAPTER 21

Record your notes and thoughts on this chapter. If you want to share these thoughts with others online, go to the bottom of the page at: http://www.interaction-design.org/encyclopedia/somaesthetics.html

**NOTES**

# Chapter 22

## Card Sorting

### *by William Hudson.*

The term card sorting applies to a wide variety of activities involving the grouping and/or naming of objects or concepts. These may be represented on physical cards; virtual cards on computer screens; or photos in either physical or computer form. Occasionally, objects themselves may be sorted. The results can be expressed in a number of ways, with the primary focus being which items were most frequently grouped together by participants and the names given to the resulting categories.

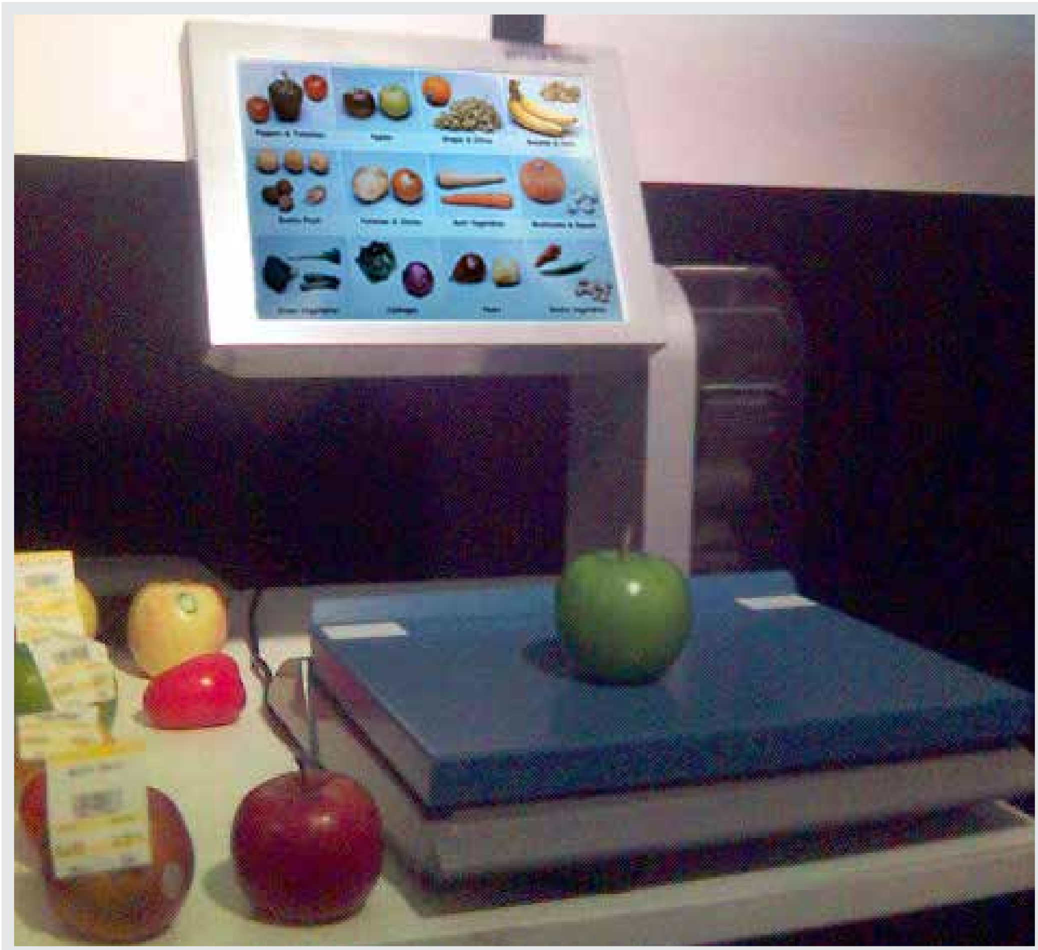
For the purpose of interaction design, the sorting process — usually performed by potential users of an interactive solution — provides:

- Terminology (what people call things)
- Relationships (proximity, similarity)
- Categories (groups and their names)

We can use this information to decide which items should be grouped together in displays; how menu contents should be organized and labelled; and perhaps most fundamentally, what words we should employ to describe the objects of our users' attention.

## 22.1 A PRACTICAL EXAMPLE

Imagine that you are responsible for the information architecture of computerized touch-screen scales of the kind increasingly common in large supermarkets, shown in Figure 22.1. The screen displays 12 images and captions at a time. There have been some complaints that customers are spending a long time at the scales and are frustrated by how the categories are organized. Table 22.1 shows a list of sample items that customers need to find. These have been printed on cards with bar codes for easy data capture (see Figure 22.2 and the Syntagm web site). Figure 22.3 shows an example of the cards organized into groups. Since this is an 'open' sort, users make up their own groups and names for them. This particular grouping represents the current solution implemented in the scales, referred to as a 'reference sort', discussed later in this chapter.



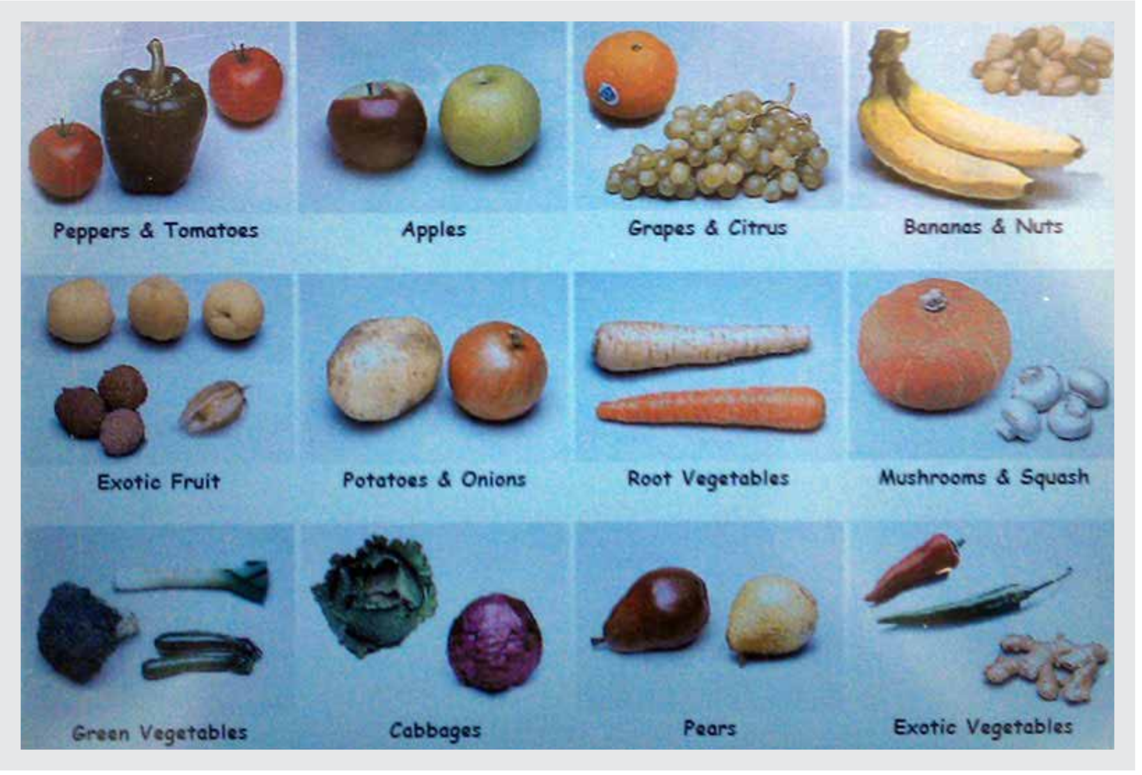





Figure 22.1 A-B: Computerized supermarket scales (touch-screen) showing categories of fruits and vegetables.

| | |
|---|---|
| Broccoli / Calabrese | Lemons |
| Carrots | Lychees |
| Chillies | Mushrooms |
| Courgettes / Zucchini | Onions |
| Fennel (bulb) | Oranges |
| Garlic | Parsnips |
| Ginger | Potatoes |
| Grapefruit | Pumpkin |
| Grapes | Squash / Marrows |
| Kiwi Fruit | Swede / Rutabaga |
| Leeks | Turnips |

Table 22.1: Sample item list of items to be found on the supermarket scales (c.f. Figure 1).

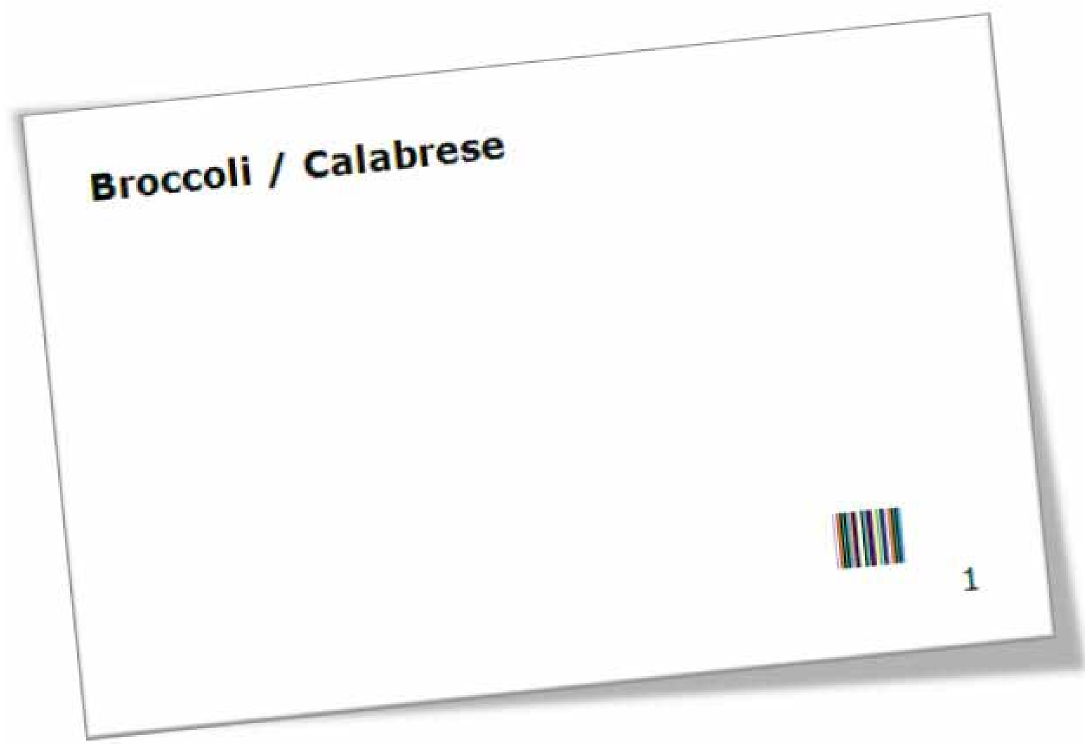



Figure 22.2: Sample card with bar codes to simplify data capture (the bar code provides the item number in machine-readable form).


Figure 22.3: Sample cards organized into groups.



Take a moment to consider how you might organize these items yourself. For most people there are at least two groups — fruit and vegetables. But in a large supermarket two groups would contain very long lists of items which would not be helpful without further subdivision. Also, there may be some terms that are unfamiliar to you. Courgette is the French name for the long, green marrow (squash) seen in British supermarkets, while zucchini is the Italian name found in the US. Conversely, what is known as a rutabaga in the US is called a swede in the UK as it was introduced to Scotland by the Swedes. Where simple language differences like these are known in advance, listing the alternatives on a single card is probably a satisfactory solution. However, in novel problem domains or in multicultural/multilingual situations where terminology is a larger issue, it may be better for participants to sort photographs or even the objects themselves (with a barcode label attached).

Whatever you are sorting, you will end up with some things (items) arranged in groups, ideally with group names. The next challenge is how to make sense of these, particularly when you have tens or hundreds of participants. No matter how the analysis is done, there are at least two things we want to know:

- What were the groups called and what was in each?
- Which items were grouped together most often?

Be careful to note that these are two separate sets of information. That grapefruit and oranges were always grouped together in the sample study is not affected by the fact that several different group names were used. Also, not surprisingly, other items were grouped with grapefruit and oranges — but the nature of these items varied with the approach taken by participants. If the group was called simply 'fruit' it contained apples, pears and other fruits as well as grapefruit and oranges. If it was called 'citrus', the only addition was lemons. So, to get a good idea of what the sort is telling us, we use different kinds of analysis. The first two correspond to the things we wanted to know:

- An **items by groups** chart shows what the groups were called and what was in each
- An **items by items** chart shows which items were grouped together most often

### 22.1.1 Items by groups chart

You can produce simple versions of the charts yourself with pencil and paper or a spreadsheet and printer. First, the items by groups chart:

- List all of the items that were sorted down the left-hand side of the page. As this needs to be done so that you can find each item quickly, alphabetic order is probably best (a word processing or spreadsheet package can help with sorting).

- Scanning through the sort results, for each new group write its name as a column heading. Place a mark in each item cell that is contained within the group. So if the first group is called ‘Citrus Fruit’, we would write this as a column heading and then mark the cells for oranges, lemons and grapefruit . Figure 22.4 shows this example.

| | Citrus Fruit |
|---|---|
| Broccoli / Calabrese | |
| Carrots | |
| Chillies | |
| Courgettes / Zucchini | |
| Fennel (bulb) | |
| Garlic | |
| Ginger | |
| Grapefruit | x |
| Grapes | |
| Kiwi Fruit | |
| Leeks | |
| Lemons | x |
| Lychees | |
| Mushrooms | |
| Onions | |
| Oranges | x |
| Parsnips | |
| Potatoes | |
| Pumpkin | |
| Squash / Marrows | |
| Swede / Rutabaga | |
| Turnips | |

Figure 22.4: Partially completed items by groups worksheet (for a single group — Citrus Fruit).

- If another participant uses the same group name (or if it is a ‘closed’ sort where you have provided all of the group names), you will only need to write the column headings once. However, for open sorts, be prepared for many variations in spelling and wording. For example, ‘soft fruit’ versus ‘berries’. It is generally best to keep such different terms separate during data capture and decide whether to merge the results at a later stage.
- If we were to reorder the items using cluster analysis (discussed later), a chart similar to that shown in Figure 22.5 would result. This has the same layout as the worksheet in Figure 22.4 — items are listed down the left-hand side and groups across the top. In the body of the chart, the square cells represent the number of times each item appeared in the named group, expressed as a shade of a chosen colour — this corresponds to the number of marks you would have made in your hand-generated version. (Percentage values are available in the application by clicking on a cell; the figure shows the result for ‘Carrots’ in the ‘Root Veg’ group.) Table 22.2 provides more details of the shading used.

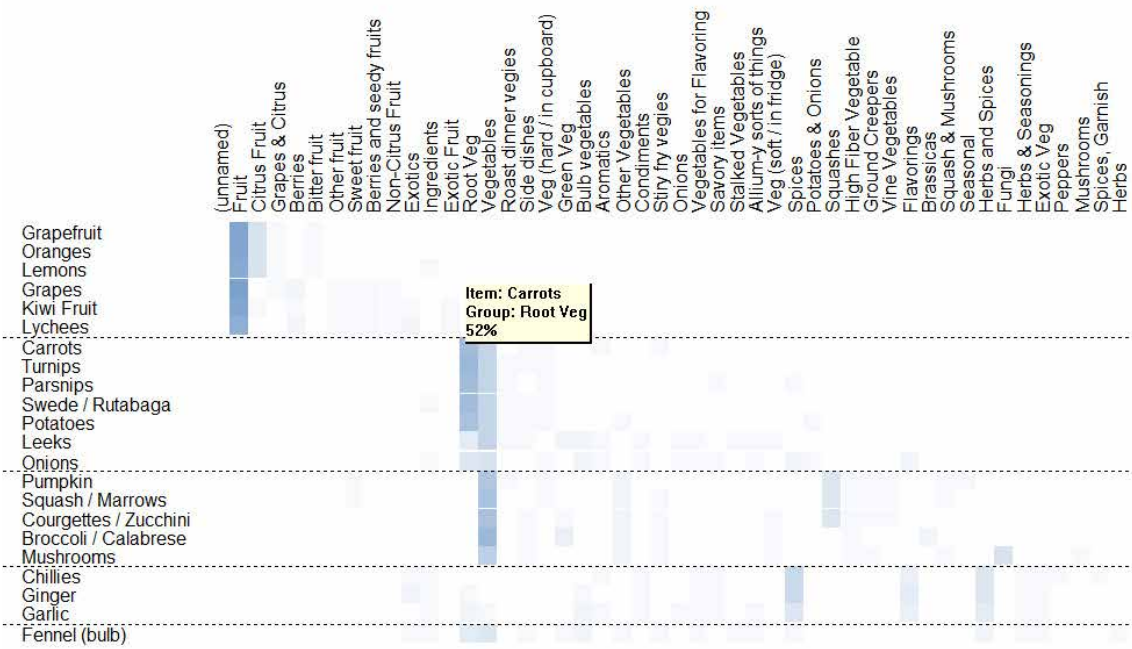



Figure 22.5: Items by groups chart of fruit & vegetables sample with 26 participants (SynCaps V2).



| | | | | | |
|---|---|---|---|---|---|
| | 0% of participants | | 25% of participants | | 50% of participants |
| | 75% of participants | | 100% of participants | | |

Table 22.2: Items by groups chart of fruit & vegetables sample with 26 participants (SynCaps V2).

### 22.1.2 Items by items chart

The items by items chart is a little more challenging to produce:

- List all of the items your participants sorted down the left-hand side of the page in alphabetic order. Repeat the list in the same order across the top of the page. You now have a matrix of items. To avoid confusion and duplicated effort, draw a line through the diagonal — from the top-left to the bottom-right, where each item meets itself, and decide which half of the matrix you are going to use. Then shade the other half. This is so you are forced to put ‘Oranges’ x ‘Grapefruit’ into the same place as ‘Grapefruit’ x ‘Oranges’. You should end up with something similar to the worksheet shown in Figure 22.6. The top-right of the matrix has been greyed-out and will not be used.

| | Broccoli / Calabrese | Carrots | Chillies | Courgettes / Zucchini | Fennel (bulb) | Garlic | Ginger | Grapefruit | Grapes | Kiwi Fruit | Leeks | Lemons | Lychees | Mushrooms | Onions | Oranges | Parsnips | Potatoes | Pumpkin | Squash / Marrows | Swede / Rutabaga | Turnips |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Broccoli / Calabrese | | | | | | | | | | | | | | | | | | | | | | |
| Carrots | | | | | | | | | | | | | | | | | | | | | | |
| Chillies | | | | | | | | | | | | | | | | | | | | | | |
| Courgettes / Zucchini | | | | | | | | | | | | | | | | | | | | | | |
| Fennel (bulb) | | | | | | | | | | | | | | | | | | | | | | |
| Garlic | | | | | | | | | | | | | | | | | | | | | | |
| Ginger | | | | | | | | | | | | | | | | | | | | | | |
| Grapefruit | | | | | | | | | | | | | | | | | | | | | | |
| Grapes | | | | | | | | | | | | | | | | | | | | | | |
| Kiwi Fruit | | | | | | | | | | | | | | | | | | | | | | |
| Leeks | | | | | | | | | | | | | | | | | | | | | | |
| Lemons | | | | | | | | | | | | | | | | | | | | | | |
| Lychees | | | | | | | | | | | | | | | | | | | | | | |
| Mushrooms | | | | | | | | | | | | | | | | | | | | | | |
| Onions | | | | | | | | | | | | | | | | | | | | | | |
| Oranges | | | | | | | | | | | | | | | | | | | | | | |
| Parsnips | | | | | | | | | | | | | | | | | | | | | | |
| Potatoes | | | | | | | | | | | | | | | | | | | | | | |
| Pumpkin | | | | | | | | | | | | | | | | | | | | | | |
| Squash / Marrows | | | | | | | | | | | | | | | | | | | | | | |
| Swede / Rutabaga | | | | | | | | | | | | | | | | | | | | | | |
| Turnips | | | | | | | | | | | | | | | | | | | | | | |

Figure 22.6: Worksheet for items by items chart (created with MS Excel).

| | Broccoli / Calabrese | Carrots | Chillies | Courgettes / Zucchini | Fennel (bulb) | Garlic | Ginger | Grapefruit | Grapes | Kiwi Fruit | Leeks | Lemons | Lychees | Mushrooms | Onions | Oranges | Parsnips | Potatoes | Pumpkin | Squash / Marrows | Swede / Rutabaga | Turnips |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Broccoli / Calabrese | | | | | | | | | | | | | | | | | | | | | | |
| Carrots | x | | | | | | | | | | | | | | | | | | | | | |
| Chillies | | | | | | | | | | | | | | | | | | | | | | |
| Courgettes / Zucchini | x | x | | | | | | | | | | | | | | | | | | | | |
| Fennel (bulb) | x | x | | x | | | | | | | | | | | | | | | | | | |
| Garlic | | | | | | | | | | | | | | | | | | | | | | |
| Ginger | | | | | | | | | | | | | | | | | | | | | | |
| Grapefruit | | | | | | | | | | | | | | | | | | | | | | |
| Grapes | | | | | | | | | | | | | | | | | | | | | | |
| Kiwi Fruit | | | | | | | | | | | | | | | | | | | | | | |
| Leeks | x | x | | x | x | | | | | | | | | | | | | | | | | |
| Lemons | | | | | | | | | | | | | | | | | | | | | | |
| Lychees | | | | | | | | | | | | | | | | | | | | | | |
| Mushrooms | | | | | | | | | | | | | | | | | | | | | | |
| Onions | x | x | | x | x | | | | | | x | | | | | | | | | | | |
| Oranges | | | | | | | | | | | | | | | | | | | | | | |
| Parsnips | x | x | | x | x | | | | | | x | | | | x | | | | | | | |
| Potatoes | x | x | | x | x | | | | | | x | | | | x | | x | | | | | |
| Pumpkin | x | x | | x | x | | | | | | x | | | | x | | x | x | | | | |
| Squash / Marrows | x | x | | x | x | | | | | | x | | | | x | | x | x | x | | | |
| Swede / Rutabaga | x | x | | x | x | | | | | | x | | | | x | | x | x | x | x | | |
| Turnips | x | x | | x | x | | | | | | x | | | | x | | x | x | x | x | x | |
| | | | | | | | | | | | | | | | | | | | | | | |
| Count | 66 | | | | | | | | | | | | | | | | | | | | | |

FIGURE 22.7: Item pairings for a group of 12 items ('Vegetables').



- Using the sorted cards, place a mark in each cell for every pair of items that appears in the same group. For example, if we came across a group called 'citrus' we would probably find it contained grapefruit, oranges and lemons, so we would mark the cells grapefruit x oranges, grapefruit x lemons and oranges x lemons. This is a simple case; for larger groups there are many marks to make: ($n^2$

— n) / 2. This is because we want all possible pairings ($n^2$) excluding items paired with themselves (-n), plus we don't need to distinguish between the order of pairs — so apples x pears is the same as pears x apples. This allows us to halve the matrix and consequently the number of marks to be made (/2 in the formula). So if you have a group of 8 items, sharpen your pencil and get ready to make marks in 28 cells. Twelve items yield 66 marks, as shown in Figure 22.7. (Bear in mind that these are the values for a single participant. Either keep a running total in each cell or add additional marks as you process subsequent participants. Alternatively, use a single sheet for each participant and simply add the results together at the end. This approach has the distinct advantage of allowing you to find and fix errors as well as making visual comparisons of participants' sorting methods.)

- Repeat for all participants. When completed, the number of marks in each cell represents how often participants grouped item pairs together. This is called an 'items by items' (or 'pairs') chart. Figure 22.8 shows a computer-generated version, with the items reordered using cluster analysis. Rather than labelling the rows and columns separately, the item names are shown on the diagonal. Note that because we have removed half the matrix, most items are folded at the diagonal. For example, 'Carrots' starts as a row on the left and then continues as a column running down the page at the diagonal. The dashed lines in the figure separate the clusters — based on the average number of groups created by participants (four).
- As for the items by groups chart (Figure 22.5), the square cells represent the percentages of participants as a shade of the selected colour as detailed in Table 22.2. In the items x items chart, however, each cell represents a pair of items that were placed together

in the same group.

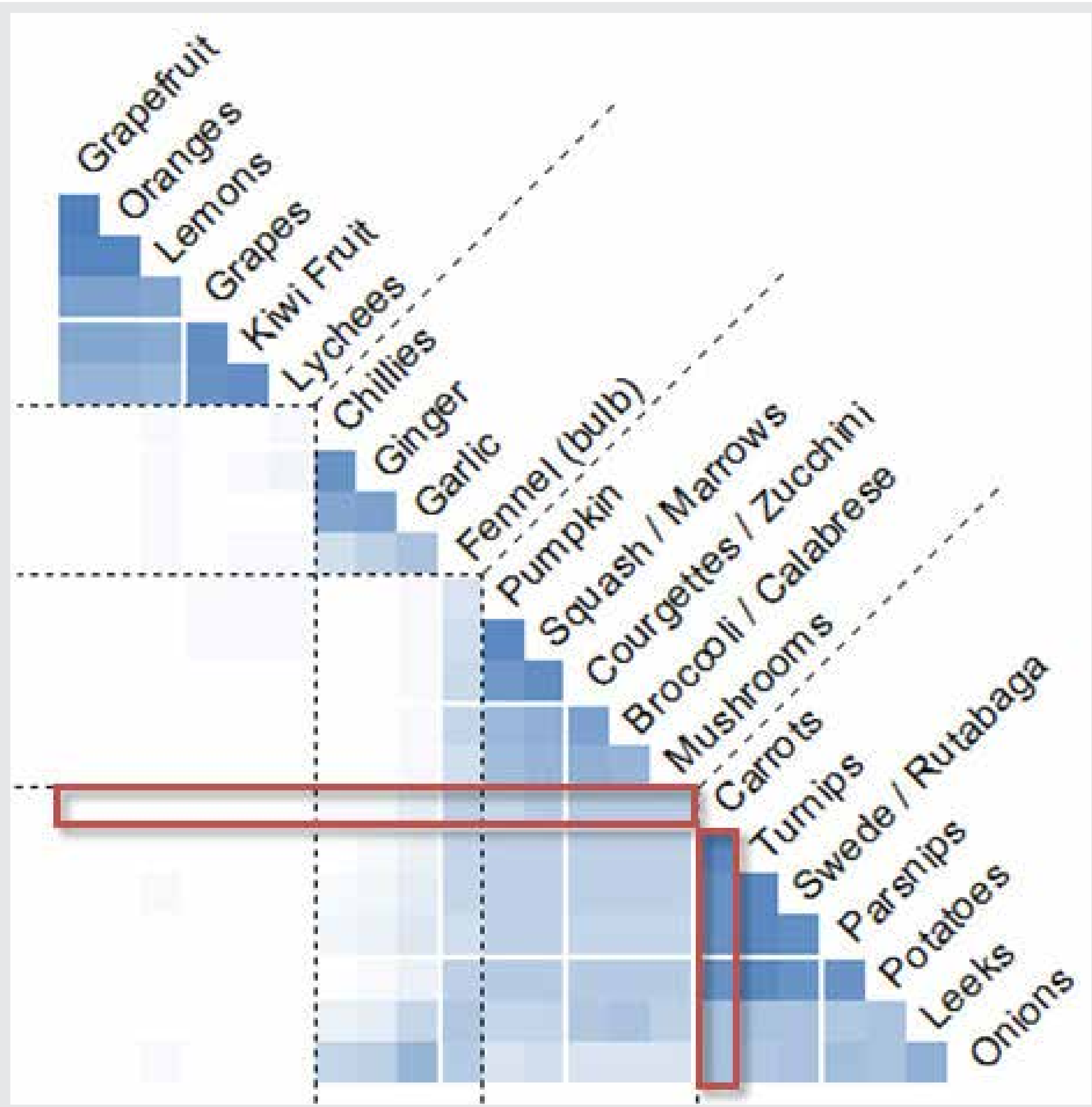



Figure 22.8: Items by items chart of fruit & vegetable sample with 26 participants (SynCaps V2) (red boxes show the complete data for 'Carrots').

## 22.2 WHAT THE ANALYSES MEAN

While it is tempting to think that a card sorting project is going to immediately provide a navigation hierarchy, this is rarely the case. The results inform a design process; they do not provide a packaged solution. The sample fruit-and-vegetable project described here provides a realistic case in point — the results are far from conclusive.

What do we know for sure from the analyses? Refer back to Figure 22.5 and Figure 22.8 and see what conclusions you can draw before proceeding.

Both charts include the results of a cluster analysis that divide the items into four groups. The items by groups chart (Figure 22.5) shows that the most popular names for the four groups were 'Fruit', 'Spices', 'Vegetables' and 'Root Veg'. 'Citrus Fruit' was a strong contender for grapefruit, oranges and lemons, while some participants (about a third) did not distinguish between 'Root veg' and 'Vegetables'.

Anything else? What about fennel? In both charts it should be possible to see that fennel has been grouped with a wide variety of other items. Although the cluster analysis placed it in the group called 'Spices' almost 20% of participants sorted it into the 'Vegetables' group. There may be nothing we can do about this other than providing access to fennel from both groups — easily done on a computerized scale or web site.

Focussing on the items by items chart for a moment, we see an important feature of the items themselves — independent of group names. Very few participants attempted to group the fruits with any of the vegetables. This shows a clear understanding of and distinction between these two main categories that we certainly should build on when designing a suitable information hierarchy. In contrast, the charts show a good deal of participant ambiguity over onions and leeks. These were frequently grouped with root vegetables, but the items by items chart shows an affinity — particularly for onions — with the group most commonly referred to as 'spices'.

What conclusions can we draw from this example? The first is that while we have learned a great deal about our participant's appreciation of the terminology, categories and concepts, the exercise was too limited for the results to be applied to a larger information space. Specifically, the small number of fruits provided in the example encouraged participants to place them in a single group. This may not be realistic in practice, although we do have some suggestions for refinement — 'Citrus Fruits' and 'Berries' from Figure 22.5. One solution would be to provide participants with a larger range of fruits, including what are called exemplars (representative types) of the categories we expect. An alternative approach would be to brief and monitor participants more closely. This is difficult to do in an online sorting activity — even if the briefing is very detailed, participants may fail to see it, read it or act on it. Most of these issues can be overcome in face-to-face sorting. If facilitators see participants producing too few categories, they can simply cajole them to create more.

So far we have touched on two popular methods of analysing card sorts — there are others which will be discussed later. But first a little background...

## 22.3 THE HISTORY OF CARD SORTING

Card sorting has a surprisingly long history, especially if the concept of categorization is included. The ancient Greeks are credited with the early development of categories, with Aristotle providing the foundations for the categorization scheme that we use today for plants and animals (Sachs 2002). The practice of sorting cards in the social sciences is somewhat more recent, but still well over 100 years old. Initially, printed playing cards were used for a variety of experiments in the nascent field of psychology (Jastrow 1886), but these were joined relatively quickly by blank cards on which researchers would write words to be categorized by subjects (Bergström 1893). Early card sorting activities were primarily concerned with establishing characteristics of the subjects — the speed of sorting used as an indicator of mental processes and reaction time (Jastrow 1886; Jastrow 1898); memory function (Bergström 1893; Bergström 1894) and imagination — using inkblots on cards (Dearborn

1898). Some of these experiments developed into what is now considered to be a standard test for neurological damage in patients who have suffered head injuries, the Wisconsin Card Sorting Test (Eling et al. 2008). In fact, card sorting was so well received in psychology that an article appeared in *Science* as early as 1914 espousing the virtues of various types of card-based activities (Kline and Kellogg 1914).

Card sorting also made its way into other fields: criminology (Galton 1891), market research (Dubois 1949), semantics (Miller 1969) and as a standard qualitative tool in the social sciences (Weller and Romney 1988; Bernard and Ryan 2009). However, it was not until the emergence of the World Wide Web in the early 1990's that card sorting was applied to the task of organizing information spaces (Nielsen and Sano 1995), with the rare exception that Tom Tullis applied card sorting to the design of menus for an operating system in the early 1980's (Tullis 1985).

### 22.3.1 Card sorting and the design of interactive products

Despite the popularity of the web, card sorting remains an under-used tool in the design of interactive products. In a survey of 217 attendees of Usability Week 2008, Nielsen Norman Group reported that the average number of card sorts conducted per year was 2. While this is twice as frequent as eye-tracking studies in the survey (average 1 per year), this is a surprisingly low number given that there are no large up-front investments required. In fact, card sorting has had only a peripheral role in interactive product design since its inception — perhaps reflecting the limited uptake of user-centred design methods in general. Peter Morville and Louis Rosenfeld devote only a few pages to card sorting in their seminal work, *Information Architecture* — now in its third edition (Morville and Rosenfeld 2006). And at the time of writing, there is only one book available on the topic of card sorting for interactive systems design, Donna Spencer's *Card Sorting: Designing Usable Categories* (Spencer 2009), which tends to be fairly conservative in terms of analysis.

## 22.4 BENEFITS OF CARD SORTING

For interaction design, customer research or research in the social sciences, few investigative techniques are as effective as card sorting in dealing with large numbers of concepts. In face-to-face settings, handling and annotating physical cards is a fairly natural and unintimidating process: observing users engaged in this process can result in many insights for researchers and provide a fertile source of questions and conversations about the problem domain being studied and, of course, users themselves. These outcomes and opportunities are hard to obtain through interviews, questionnaires and usability evaluations, although each of these alternatives has its strengths for more limited scopes of investigation. For example, it is relatively easy to discover that a single menu item is mislabelled in a usability study, but prohibitively expensive for several dozen items.

## 22.5 QUALITATIVE VERSUS QUANTITATIVE OUTCOMES

At one extreme, card sorts can be conducted on a one-to-one basis as a tool for discovery (knowledge elicitation) and a means of generating meaningful discussion between participants and researchers (Weller and Romney 1988; Bernard and Ryan 2009). The outcomes here are generally a better understanding of the problem domain from a user's perspective with terms, relationships and categories expressed in the resulting groups. At the other extreme, it is very easy to organize online sorts with hundreds of participants to discover whether the terminology and concepts presented are well understood across a large user population (Fincher and Tenenberg 2005). While results in the one-to-one approach are primarily qualitative, those of the large-scale online studies are mostly quantitative. (Note that it is not impossible to obtain qualitative information from online studies; there simply are not as many opportunities to persuade or allow online participants to provide useful feedback.)

## 22.6 WHAT TO SORT

Not surprisingly, the choice of what to have participants sort depends largely on what a researcher, information architect or interaction designer is trying to discover. For 'green-field' projects — those that lack any constraints imposed by prior work — a first priority would be to establish a vocabulary. In this context, users could be presented with objects, images or descriptions of items and asked to name them. Once named, they could be grouped, with the groups in turn also named. This is fairly easy to do in face-to-face settings, where numbered or bar-coded labels can be applied to objects or photos (see, for example, the card sorting templates for Microsoft Word at the Syntagm web site). Note that some web-based sorting packages, such as websort.net, do allow photos to be sorted, but provide no means for users to apply names to the items depicted.

- **Fixed Items:** If terminology is already established and immutable (such as product names), then basic research as described above is unnecessary. The primary goal of a sorting activity would be to discover which items should be grouped together and what these groups should be called. This is a relatively straightforward undertaking for either face-to-face or online approaches. The choice would largely be determined by whether qualitative feedback is desired (for which face-to-face sorting with paper cards would be most appropriate) or if qualitative feedback using larger numbers of participants would be beneficial. Good quality results can be obtained from 15-30 participants in a face-to-face context (Nielsen 2004; Tullis and Wood 2004) while online sorts can be conducted for hundreds of participants at no additional cost except for recruitment. Also, large-scale studies can be useful for increasing engagement within an organization or ensuring that a diverse collection of users have a similar understanding of a problem domain.

- **User Goals:** Card sorting is frequently applied to navigation design. However, simply listing the names of documents, pages or features that will be present in a solution does not guarantee that users will be able to reach their goals, even if they are organized optimally. Starting with user goals helps to ensure that navigation design is effective. So rather than asking participants to sort items such as “Employee Manual”, “Staff Policies” and “HR Guide” (all of which confusingly overlap), consider instead the goals that users have in accessing these documents: “Find holiday entitlement”, “Can I work at home?”, “How much time can I take off for a new baby?” and so on. (Tom Tullis employed user goals in his design of operating system menus — (See Tullis 1985)). Server logs, particularly search phrases; content audits; and user research can be used to build a list of user goals, with card sorting providing grouping and category names.
- **Multilevel Hierarchies:** Most sorting and analysis tools do not support the kind of multiple-level hierarchies found in all but the simplest interactive solutions. Even the produce scales used in the sample card sort could use a multilevel hierarchy. For example, a top-level category called ‘Fruit’ might lead to ‘Citrus Fruit’, ‘Apples and Pears’, ‘Exotic Fruit’ and so on. However, the lack of analysis support for multilevel hierarchies is not an insurmountable problem. In fact, multilevel hierarchies at the analysis stage can increase the complexity of a sorting activity substantially, thereby making it a daunting undertaking for many participants. Instead, conduct multiple single-level sorting activities. Focus on the lowest levels

(the ‘leaves’ of the navigation tree) since category names provided by participants often vary considerably in their levels of abstraction, as we saw in the example. Participants’ category names included ‘Fruit’, ‘Soft fruit’ and ‘Berries’. Each of these could be appropriate for higher-level navigation headings. (Multilevel sorting is discussed in more detail under *Section 22.9, Advanced analysis*.)

## 22.7 HOW TO DO A CARD SORT


Markus
Chris

### 22.7.1 Choosing an approach

Face-to-face sorting methods are generally better for qualitative research, while online methods (web-based or desktop) are more appropriate for quantitative results. However, this is not always true; for example, it would be possible to sit with a participant or share their desktop while they conducted an online sort. This could result in good qualitative data, but it would be more intimidating for participants and much harder work for the facilitator. Remote desktop sharing can also be technically challenging, especially in the presence of corporate firewalls and security policies.

**Researchers or interaction designers can also choose between**

- open sorting, where users make up their own categories
- closed sorting, where categories are predefined
- hybrid sorting; some combination of the two

For most purposes, open sorting is the best choice, although supplying some predefined categories is always helpful to participants and is supported by most sorting and analysis tools. Closed sorting can be used when trying to establish changes required to an existing structure, particularly with analysis tools that provide comparisons between a 'reference sort' (such as an existing or proposed solution) and participants' results — see Figure 22.9.

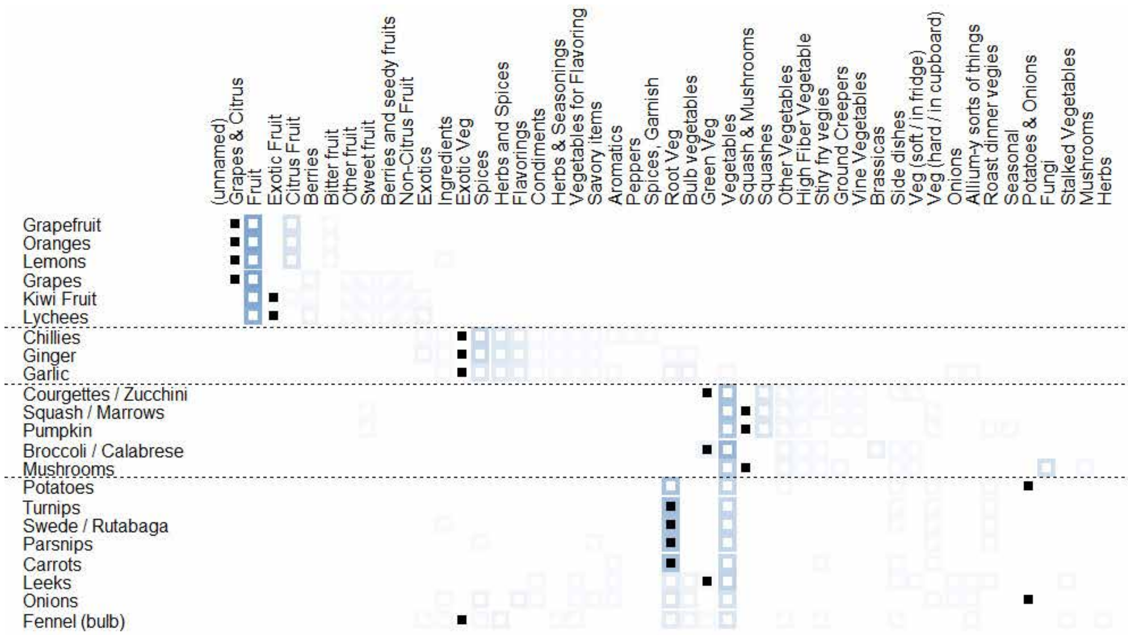



Figure 22.9: Fruit & vegetable example with reference sort showing existing solution (SynCaps V2).



In this items by groups chart, Figure 22.9, the current solution is shown with black squares in a cell. So while most participants choose to group all fruit together, the computerized scales used two unusual groups; 'Grapes & Citrus' and 'Exotic Fruit'. However, there were some areas of correspondence: many participants agreed with the current design for the 'root veg' group towards the bottom-centre of the chart.

### 22.7.2 Recruiting and briefing participants

As with any other form of user-centred design, participants of a card sorting activity should be representative of the users envisaged for the solution. However, given the difficulties that some members of the population may have with technology (older users for example) it is often beneficial to over-sample these groups to ensure that the resulting design is effective for as broad an audience as possible. Where possible, try to use participants who are motivated to participate by interests that are more than purely monetary — existing users or customers for example.

When briefing participants for a sorting activity, it does not pay to be too vague in stating the requirements. In navigation design the number of categories needed for a set of items is not a complete mystery. There is usually a balance to be struck between the number and size of groups (Kiger 1984). Consequently, it is important to provide participants with adequate information about the number and level of groups you require. If you are trying to devise menus for our computerised produce scale having space for 12 items on the screen, do not be shy about letting participants know that. Similarly, horizontal menu bars on websites or desktop applications rarely have space for more than 6 or 8 items. Allowing participants to generate 20 or 30 categories in these cases is potentially a waste of their time and yours.

Similarly, if you have group names that you know, or at least strongly suspect you need, provide those to participants. This can be done in both face-to-face and online settings. But do encourage participants to make up their own group names if they prefer.

Participants should also be advised on how to deal with items they do not understand. While some researchers or interaction designers suggest that all items should be sorted — leaving participants simply to guess at those they do not recognize — this can lead to spurious groupings. Consider asking users simply not to sort items they do not recognize, or create a specific 'unknown' group to receive them. These can then be excluded from the results. Most online sorting tools now do allow items to remain unsorted. However, make sure that analysis results are based on the number of participants rather than the number of times that an item was sorted.

### 22.7.3 Time to sort

The amount of time required to perform a sort can vary considerably from person to person, but is largely dependent on the number of items to be sorted:

- Approximately 20 minutes for 30 items
- 30 minutes for 50 items
- 60 minutes for 100 items

However, other factors include how familiar the terms and concepts are to participants and how motivated they are to provide results conscientiously. Also, it is possible to sort up to 150 cards in single sessions, but higher quality results might be obtained by splitting such a large project into smaller parts.

### 22.7.4 Preparing a sort

For face-to-face (paper-based) sorting, getting items and group names onto cards can be a tedious undertaking. Happily, standard mail-merge software can be used to make this task easier, meaning that items can be printed either directly onto cards or self-adhesive labels. Free mail-merge templates for Microsoft Word can be found on the Syntagm web site for both North American and European paper sizes. These also include bar codes that can be used to simplify data collection: instead of typing in an item name or number, the bar codes allow them to be read directly using a simple USB scanner. This is both quicker and less error-prone than manual entry — it makes it relatively easy to process 120 cards or more per minute (full instructions are included on the web page referred to).

Preparation for online sorting is relatively straightforward, requiring only lists of the items and group names (if any) to be uploaded.

However, regardless of the method of sorting, be aware that superficial similarities in the names used can produce unhelpful results. Consider these menu item names from an intranet:

- **Manage** absence and holidays
- **Manage** difficult colleagues
- Change **management**

If faced with a large number of items to sort, participants may simply group similar names together. This is called a superficial match. To overcome this, consider modifying the item names:

- Absence and holidays
- Coping with difficult colleagues
- Change management

In the first two items the word 'manage' was not an essential part of the name. Removing it or using a synonym prevents unwanted grouping.

### 22.7.5 Choosing names

Apart from the issue of superficial similarities mentioned above, be careful to choose names that are in common use, especially where interactive solutions are being designed for a broad range of abilities. This is not just common sense, but also a requirement of disability discrimination legislation in many countries. Put simply, language should be no more complex than needed to convey the required information. In English, longer words (measured in syllables) are used much less frequently than shorter ones (Klare 1963). And even though participants in a card sort might suggest unusual names for items or groups — such as 'brassicas' — most people will go into their local supermarket or green grocers asking for cabbage rather than use its Latin genus. If in doubt, consult a reference on common words such as the Corpus of Contemporary American English, the British National Corpus or similar sources for other languages.

## 22.8 HOW TO UNDERSTAND THE RESULTS

For very small projects, just leafing through the sorted cards or listing of online results can provide useful insights into groupings. However, larger projects will require some form of analysis, ranging from simple tabulation through to cluster analysis. Note that while cluster analysis is potentially a very complex subject (Romesburg 2004; Bernard and Ryan 2009), most card sorting tools use a fairly simple form of cluster analysis that could easily be replicated manually. It is known as 'hierarchical cluster analysis'. The 'hierarchy' in this case refers to the way in which smaller clusters are aggregated to form larger ones until all are included.

### 22.8.1 Simple analysis

Simple tabulation of items by groups can be performed manually (as described above) or by using a spreadsheet package such as can be found at Boxes and Arrows. However, online sorting tools will do this analysis for you. For printed cards using the Microsoft Word mail-merge templates described earlier, SynCaps V2 and later will produce items by items, items by groups and dendrogram analyses.

| | Hiring New People | Joining | Learning & Development | Managing People | Pay & Benefits | Policies and procedures | Time Off | Work& Life Changes | Your Career at |
|---|---|---|---|---|---|---|---|---|---|
| Hiring New People | 92% | 3% | 0% | 2% | | 1% | 0% | | 0% |
| Find People to Hire | 90% | 4% | 0% | 2% | 0% | 1% | 0% | 0% | 1% |
| Choosing New People | 88% | 4% | 0% | 5% | 0% | 1% | 0% | 0% | 0% |
| Recommend a Friend | 58% | 19% | | 0% | 5% | 9% | 0% | 1% | 5% |
| Requesting an Employee Reference | 31% | 6% | 0% | 14% | 2% | 28% | | 4% | 13% |
| Getting Started at | 6% | 84% | 1% | 0% | 0% | 1% | 0% | 0% | 6% |
| Getting a New Person Started | 39% | 46% | 1% | 9% | 1% | 1% | 0% | 0% | 1% |
| Get Information for New Team Members | 25% | 41% | 5% | 18% | 0% | 7% | 0% | 1% | 2% |
| About Learning and Training | 0% | 1% | 93% | 0% | | 1% | | 0% | 2% |
| Acquiring New Knowledge | 0% | 0% | 92% | 0% | | 0% | 0% | 0% | 5% |
| Learning New Skills | | 1% | 92% | 0% | | 0% | 0% | 0% | 5% |
| Your Required Learning | 0% | 2% | 88% | 1% | 0% | 1% | | 0% | 6% |
| Identify and Develop Your Natural Abilities | | 0% | 76% | 3% | 0% | | 0% | 0% | 18% |
| Helping People Develop | 0% | 0% | 58% | 35% | | 0% | 0% | 1% | 4% |
| Learning from Formal Reviews | 1% | 0% | 54% | 24% | 0% | 2% | | 1% | 16% |
| Learning from Informal Reviews | 0% | | 54% | 26% | 0% | 2% | 0% | 1% | 15% |
| Using World | 0% | 26% | 32% | 1% | 0% | 22% | 1% | 4% | 14% |
| Managing Performance Problems | | | 3% | 90% | 0% | 4% | 0% | 0% | 1% |
| Managing Team Member Complaints | 0% | | 0% | 88% | 0% | 9% | | 0% | 0% |
| Managing Work Behaviour | | | 3% | 86% | 0% | 3% | 0% | 3% | 2% |
| Managing People's Skills | | | 10% | 86% | 0% | 1% | 0% | 0% | 1% |
| Keeping Good People | 2% | 1% | 3% | 79% | 1% | 2% | 0% | 1% | 10% |
| Holding Informal Reviews | 1% | 0% | 7% | 78% | | 6% | 0% | 1% | 5% |
| Conducting Formal Reviews | 1% | 0% | 6% | 77% | 1% | 8% | 0% | 1% | 4% |
| Dealing with Unacceptable Behaviour | 0% | 0% | 1% | 76% | 0% | 21% | 0% | 0% | 0% |
| Understanding People's Potential | 2% | | 20% | 69% | 0% | 0% | 0% | 0% | 7% |
| Dealing with Inadequate Skills | 0% | | 37% | 56% | 0% | 3% | | 1% | 1% |
| Dealing with Team Member Job Changes | 1% | 0% | 2% | 56% | 0% | 5% | 0% | 31% | 3% |
| Working Together Effectively | | 4% | 20% | 44% | | 6% | 0% | 7% | 17% |
| Providing References for Leavers | 1% | 0% | 0% | 37% | 1% | 36% | 2% | 11% | 9% |
| How We Work Together | | 13% | 15% | 27% | 0% | 14% | 0% | 9% | 19% |

FIGURE 22.10: Items by groups chart for intranet navigation created using the web-based optimal sort software.



Figure 22.10 is an items by groups chart showing an alternative presentation to that of Figure 22.5. In both cases, the items are listed down the left-hand side of the chart with the group names across the top. A cluster analysis has been performed to determine which items are most closely related, producing an item ordering that moves from one cluster to the next. The only significant difference between the two figures is that Figure 22.5 uses shading to show the relative strength of each relationship (figures are available by clicking on a cell) while Figure 22.10 presents the percentage figures with blue shading only to highlight the most significant results.

### 22.8.2 Cluster analysis

The type of cluster analysis performed by most card sorting tools is 'hierarchical cluster analysis' or HCA. The usual result is a graphical display called a dendrogram, or sometimes 'dendrogram', which has its roots (literally) in the Greek word for 'tree', which is 'dendron'.

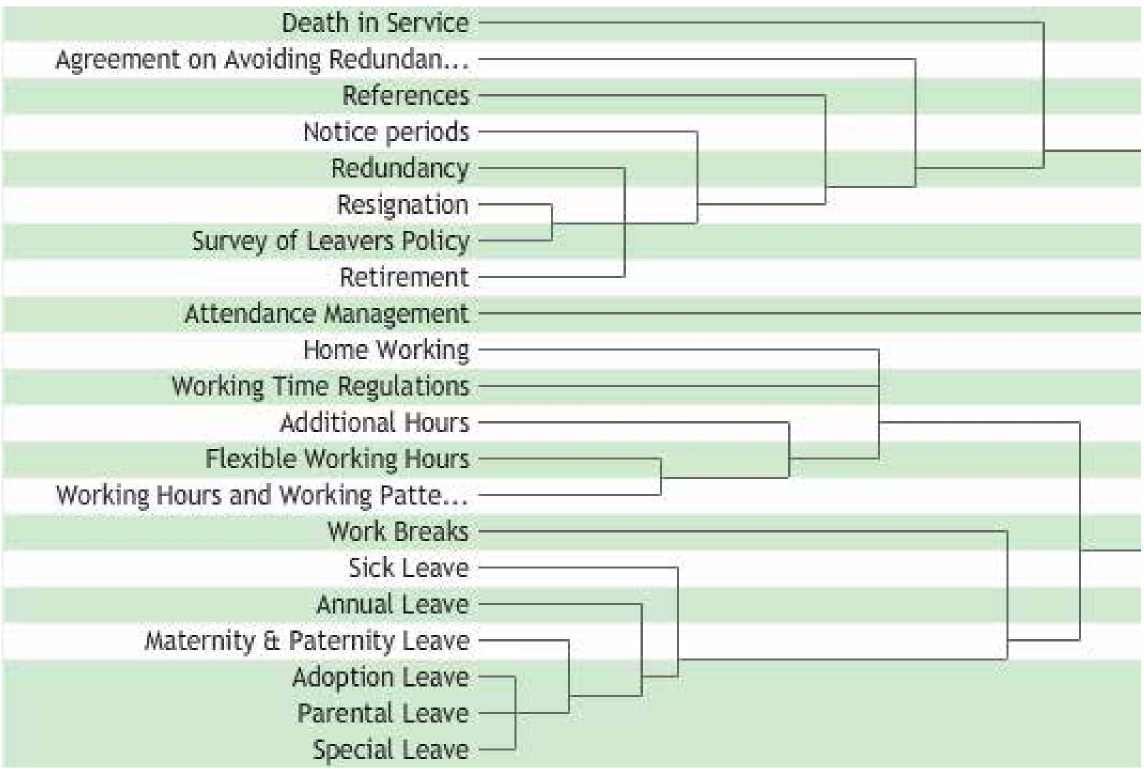



Figure 22.11: Dendrogram of Intranet Navigation created using the web-based Websort software.



Figure 22.11 shows a hierarchical cluster analysis in the form of a dendrogram. The example is taken from an intranet navigation sorting activity. The hierarchical nature of the dendrogram is related to the strength of the relationships between items, as measured by how frequently they appeared in the same groups. And as in real trees, shorter branches are stronger. In Figure 22.11, the six items at the bottom all include the word ‘leave’. However, participants have primarily grouped ‘Adoption -’, ‘Parental -’ and ‘Special’ Leave as being closely related, but were less consistent with ‘Maternity & Paternity - ‘, ‘Annual -’ and ‘Sick’ Leave. Finally, ‘Work Breaks’ was sometimes grouped with the leave items, but this relationship is fairly weak compared with the others. If you wanted to know *why* the work breaks item relationship is weaker, you would need to consult an items by items chart, an items by groups chart or the raw proximity matrix if available — the latter simply showing the number of times each pairing of items appeared together in the same groups.

The dendrogram also gives some insight to the way the cluster analysis works. The method used is called ‘agglomerative clustering’, meaning simply that we build the clusters from the bottom up. So in the intranet example, the first cluster would have started with the last three ‘leave’ items — they have the shortest branches — with ‘Maternity & Paternity Leave’ subsequently subsumed. Then, looking again at Figure 22.11, the next strongest relationship appears towards the top of the chart, as ‘Resignation’ and ‘Survey of leavers policy’.

As items are agglomerated into clusters, an average score (based again on the number of times pairs of items appeared in the same groups) is calculated. This is shown in the dendrogram by how far the vertical connecting lines are from the labels. As mentioned above, the resulting branches reflect stronger relationships when they are shorter — that is, when the vertical connecting lines are closer to the labels, as for the bottom three items in Figure 22.11.

In a dendrogram, clusters are joined together into branches until all items have been included. This means that the weakest relationships — between dissimilar clusters — can be found furthest from the item labels. Although Figure 22.11 does not show a complete dendrogram, it does include three long branches that are continued off to the right. These represent three dissimilar clusters; each will require their own category labels (which could be derived from an items by groups chart). Note that dendrograms take no account of group names; it may well be that even though ‘Adoption -’, ‘Parental -’ and ‘Special Leave’ were grouped together frequently, participants may have applied a wide variety of names to that grouping. Also be aware that in a dendrogram, items can appear in only one place. Therefore, if an item was split equally by participants between two different groups, it would appear only as a weak relationship in one of them. You would need to visit the items by groups chart to notice this.

## 22.9 ADVANCED ANALYSIS

In trying to make sense of card sorting results, there are two problems that frequently recur. The first is that not all participants have the same motivation, experience or needs. This means that we may have participants whose sort results are simply ‘noise’ — particularly for online sorts with an attractive incentive. In other cases we may believe we have one relatively homogenous group of participants, when in fact we have multiple. This can be due to general factors such as experience — in which case we need to accommodate these multiple groups in our designs; or it may be due to different contexts of use. In the latter case we should try to understand the differences and to decide whether separate designs are warranted. Unfortunately, traditional card sorting analysis tools are not much help here. But some of this information can be obtained manually — by examining the number and size of groups produced by each participant, for example: those in a hurry tend to have fewer groups and a large number of items in unhelpful categories such as ‘don’t know’ or ‘miscellaneous’, while those who have a substantially different view of the problem domain may produce an unusual number of groups (relative to the average). Optimal Workshop has added some participant-oriented results to their web-based service. Fairly detailed participant and item spreadsheets can be found in all versions of SynCaps.

The second recurring problem is related to the basic principle of cluster analysis: every item is assigned to exactly one cluster. To a certain extent, this can be worked-around by careful inspection of the items by items and items by groups analyses. For example, an item such as a cucumber might be split equally between 'green vegetables' and 'salad vegetables'. It will appear in the dendrogram in either of these groups — the choice will be arbitrary if the split is exactly 50:50 — with a fairly weak relationship. However, the weakness of the relationship is not because participants were confused about where it should go; they just did not agree. The items by items and items by groups charts would show this clearly. However, because of this limitation of cluster analysis, some researchers have explored other advanced statistical techniques; most notably factor analysis. See Capra 2005 and Giovannini 2012. A more detailed account of card sorting analysis methods can be found in (Corter 1996) and (Coxon 1999).

### 22.9.1 Multilevel sorting

The primary method of sorting discussed in this chapter can be described as single-level or 'flat'. Participants are given a set of items which they should sort into a single level of groups. So while it might be tempting to nest groups — 'leaf vegetables' within 'green vegetables' within 'vegetables', for example — there are two issues to be aware of:

- Limitations of analysis: The most common methods of analysis use a single measure of closeness or proximity of related items. This is based on how frequently items were placed together by participants. It is not practical to perform a cluster analysis on multiple group levels, but it is relatively straightforward to apply weightings to item proximities according to whether they appeared in the same group, a sub-group, a sub-sub-group and so on. Items that appear together in the same group would receive the highest weighting, pairs split between immediate subgroups a slightly lower weighting; and those split between second-order sub-groups lower still (and so on). For example, cucumber and courgette/zucchini would receive the maximum weighting if they both appeared in a ‘Green Vegetables’ group but a lower weighting if courgettes/ zucchini appeared in a group named ‘Green Vegetables’ and cucumber in a sub-group named ‘Salad Vegetables’ (illustrated in Figure 22.12 using a maximum weighting of 2). This is the approach taken by the (now defunct) EZsort/Usort (Dong et al. 2001) and the free SynCaps V1 packages in their anonymous single-level sub-groups implementation. (Anonymous sub-groups are simply unnamed.) This has been extended to multiple levels by packages such as UXsort (uxsort.com) and SynCaps V3 (Syntagm Ltd). SynCaps V3 also provides an analysis of sub-group names used at each level. See Harloff 2005 for a further discussion of weighted multi-level sorts.

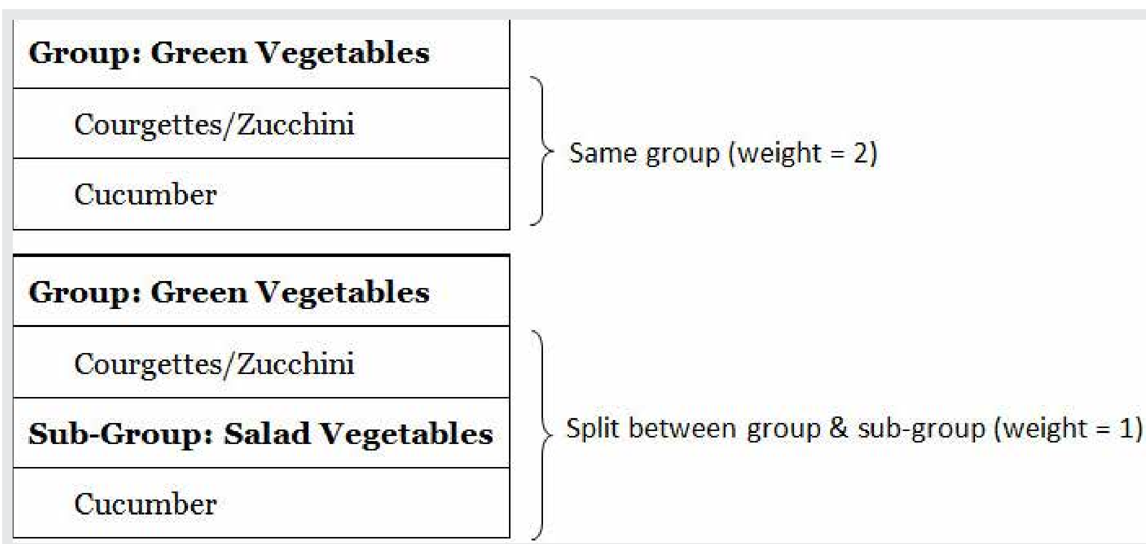



Figure 22.12: Example of multilevel weighting.

- Scale and complexity: One of the biggest challenges with multilevel card sorting is the considerable increase in the number of items to be sorted and the resulting solutions (Wood and Wood 2008). Consequently, it would be inadvisable to give participants the entire navigation hierarchy of a large intranet or e-commerce site and ask them to organize these as they see fit. Participants in card sorts are users, not information architects. Multilevel card sorting is much more likely to be effective when the potential solutions are partially defined or constrained. Even then, researchers and designers may get more useful information from a series of single-level sorting activities where this is practical.

## 22.10 TREE SORTING

Tree sorting (also called 'tree testing' and 'reverse card sorting') is a concept related to card sorting, but in many respects quite different. In essence it is a simulation of a navigation tree that would be found in a software application or web site. Online participants are presented with goals and then asked to navigate using the tree simulation. Figure 22.13 illustrates the process across several screens (step 1 is the first screen; step 2 is the second and so on). In step 1, the participant has chosen 'Fruit', while in step 2 'Soft Fruit' was selected. If the wrong selections are made, participants will need to back-track to find a more appropriate menu. A large number of tasks can be made available, with only a random subset displayed to each participant if required.

On completion of a project, researchers and designers can be presented with success rates, error rates and time taken (or related variations). While closed card sorting can be of some help in validating a navigation design, tree sorting is a more effective approach in most cases. (See plainframe.com and optimalworkshop.com)


Figure 22.13: Tree sorting example with a task of 'find oranges' from optimalwork-shop.com.



## 22.11 WHERE TO LEARN MORE

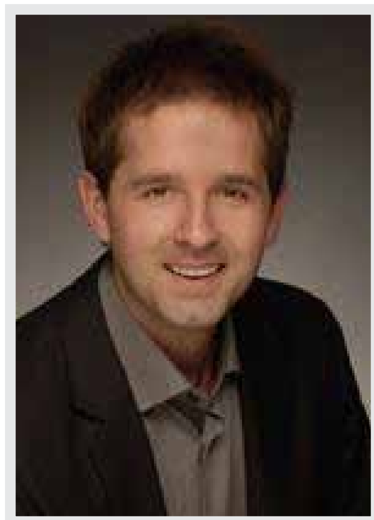
Aside from the references listed below and particularly Donna Spencer's on card sorting (Spencer 2009), there are a number of helpful web resources:

- Card Sorting (Usability.gov)
- Card Sorting and Computer-Aided Paper Sorting (Syntagm Ltd)

## 22.12 COMMENTARY BY JEFF SAURO

How to cite this commentary in your report

### Jeff Sauro



*© Jeff Sauro*

Jeff is a Six-Sigma trained statistical analyst and pioneer in quantifying the user experience. He is founding principal of Measuring Usability LLC, a quantitative user research firm based in Denver, CO. He is author of four books including: Quantifying the User Experience: Practical Statistics for User Research. He has worked for GE, Intuit, PeopleSoft and Oracle and has consulted with...

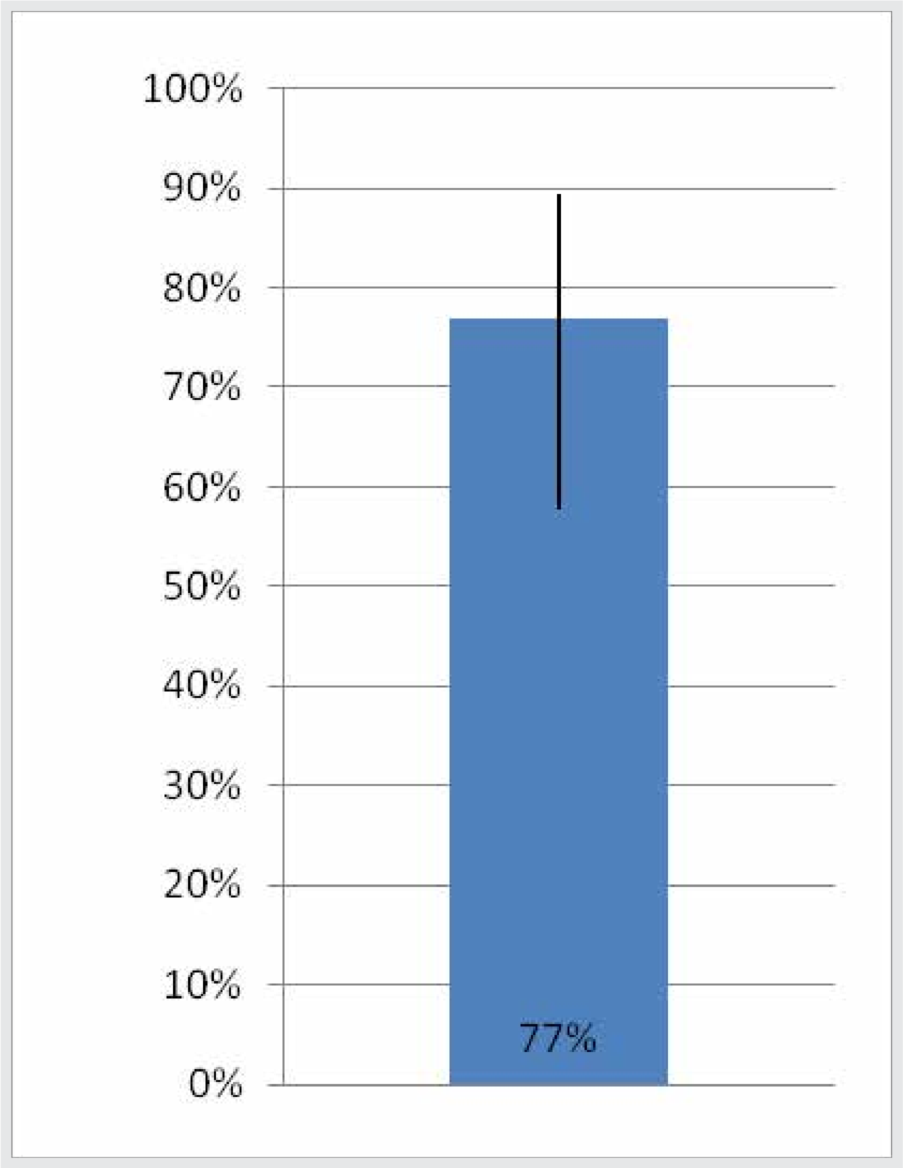
Jeff Sauro
Jeff Sauro is a member of The Interaction Design Foundation

What is it, how do we use it, where did it come from and how do we interpret the results? That's what you want to know when using a method like card sorting. Hudson delivers succinct points and comprehensive coverage on this essential UX method. He accurately articulates how card sorting generates both qualitative and quantitative data and illustrates how interpreting one of the signature graphs of card sorting (the dendrogram) involves both data and judgment. Here are a few more points to consider when quantifying the results of a card sort.

### 22.12.1 Confidence Intervals

Card sorts - like most User Research methods - involve working with a sample (often a small one) of the larger user population. With any sample comes uncertainty as to how stable the numbers are. One of the most effective strategies is to add confidence intervals around the sample statistics. A confidence interval tells us the most plausible range for the unknown population percentages.

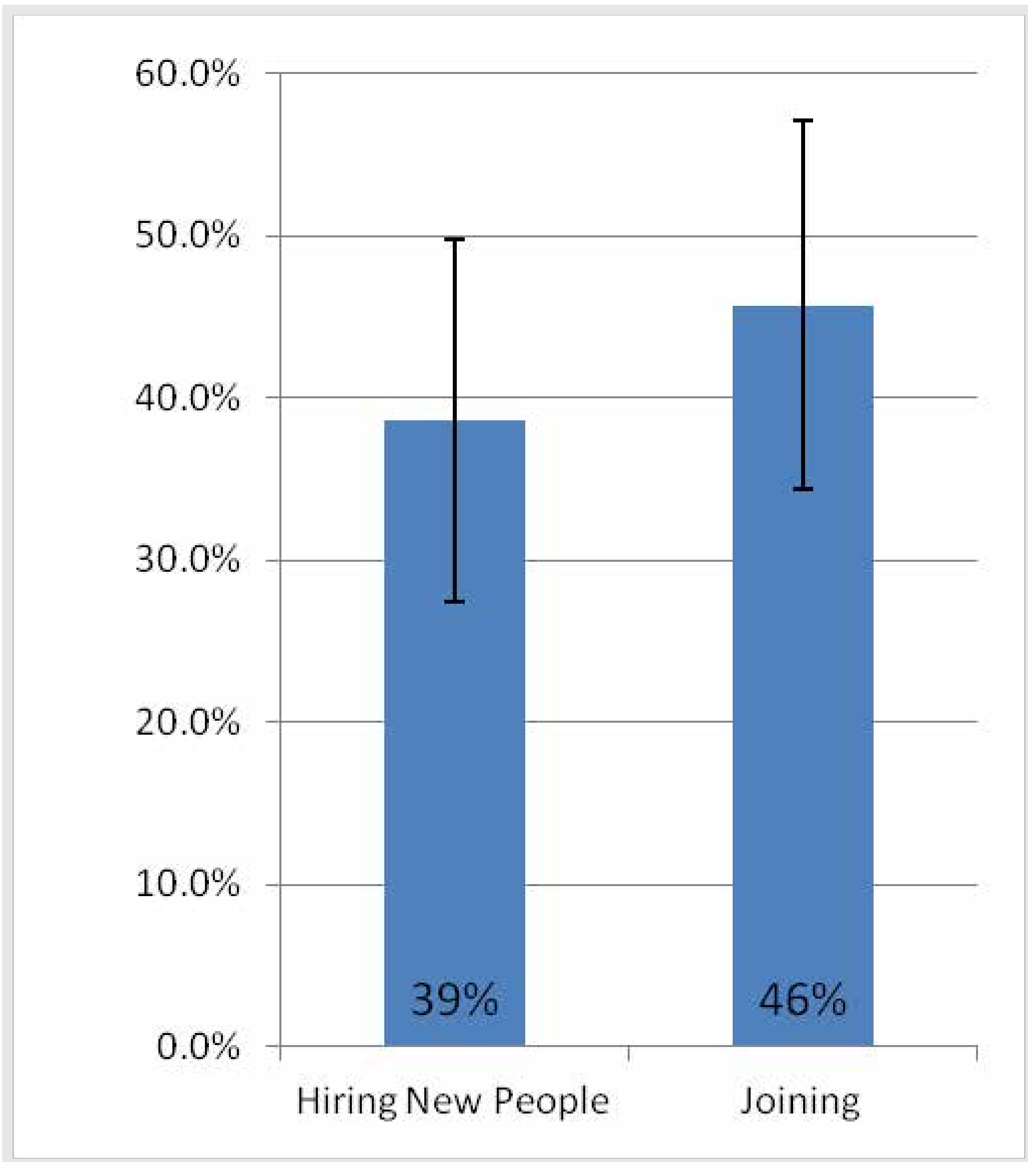
For example, let's assume 20 out of 26 users (77%) were able to successfully find Strawberries under the "Soft Fruit" category (Figure 22.13). Even without measuring all users, we can then be 95% confident between 58% and 89% of all users would successfully locate strawberries (assuming our sample is reasonably representative).






The margin of error around our percentage is +/- ~16%. The lower boundary of the confidence interval tells us that we can be 95% confident 58% **or more** of users would find the location of Strawberries. If we have as a rudimentary goal to have most users find the fruit then we have evidence of achieving this goal.

We can apply the same method to qualifying the percentages of cards placed into a category. For example, let's assume 70 participants conducted the card-sort shown in Figure 22.10. We see that 46% placed "Getting a New Person Started" in the "Joining" category but 39% placed this card in the "Hiring New People" category.



The 95% confidence interval for the "Hiring New People" category is between 35% and 58% and between 28% and 50% for the "Joining" category (see the figure above). The substantial overlap in the confidence intervals means we shouldn't have much confidence in this difference. An online calculator is available at http://www.measuringusability.com/wald.htm to make the computations.

Due to the large overlap in the intervals we cannot distinguish the 5 percentage point difference from sampling error. If we need to pick one we should go with "Joining" but we should consider both categories as viable options.

### 22.12.2 Sample Sizes

As with most evaluations, when involving users one of the first questions asked is "How many users do I need?" Surprisingly, there is little guidance on determining your sample size other than the 2004 Tullis and Wood article. Tullis and Wood performed a resampling study with one large card sort involving 168 users and found the cluster results would have been very similar (correlations above .93) at sample sizes between 20-30.

This sample size is based on the particulars of a single study (agreement in card placement and 46 cards) and on viewing the dendrogram so the results are most appropriate if your study is similar to theirs.

Another approach to sample size planning is based on the percent of users who place cards in each category chart (Figure 22.10) or correctly select the right path in tree testing (Figure 22.13). This approach is based on working backwards from the confidence intervals like those generated in the previous section. In the first example we had a margin of error of 16% around the percent of users who would correctly locate strawberries.

If we wanted to generate a more precise estimate, and cut our margin of error in half to +/- 8% we work backwards from the confidence interval and get a required sample size of 147. The following table shows the expected margins of error for 95% confidence intervals at different sample sizes.

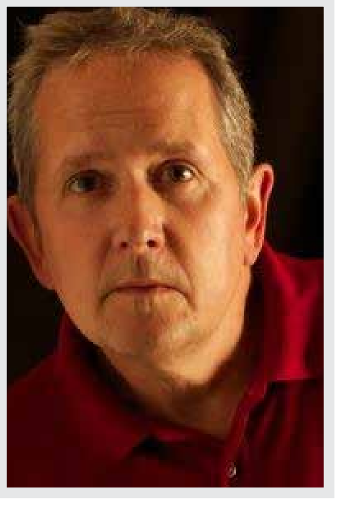
| Sample Size | Margin of Error (+/-) |
| --- | --- |
| 10 | 27% |
| 21 | 20% |
| 30 | 17% |
| 39 | 15% |
| 53 | 13% |
| 93 | 10% |
| 115 | 9% |
| 147 | 8% |
| 193 | 7% |
| 263 | 6% |
| 381 | 5% |
| 597 | 4% |
| 1064 | 3% |

The computations are explained in Chapter 3 and Chapter 6 of Quantifying the User Experience.

## 22.13 COMMENTARY BY DAVID TRAVIS

How to cite this commentary in your report

### David Travis



© *David Travis*

David Travis is the founder of Userfocus. He holds a BSc (Hons) degree and a PhD in Psychology and he is a Chartered Psychologist. His professional affiliations include membership of the Experimental Psychology Society, the Information Architecture Institute and the Usability Professionals Association and he is an Associate Fellow of the British Psychological Society. His career spans th...

David Travis
David Travis is a member of The Interaction Design Foundation

William Hudson writes knowledgeably and expertly about card sorting — as you would expect from someone who has been practising the technique for well over a decade. William's chapter in the encyclopaedia will be a great help to those people new to card sorting who need a step-by-step tutorial through the technique.

For people who already have some experience with card sorting, I wanted to add a few words about dealing with some of the problems that come up when you do open and closed card sorting in practice. First: with an open card sort, how do you deal with a very large web site where you may have hundreds of items that need to be sorted? And second: with a closed card sort, how can you present the results back to clients in such a way that they understand the complex, quantitative data you have collected?

### 22.13.1 An open card sort with a very large web site

A few years ago, I worked with an auction web site to help them revise their online help system. There was a large number of help pages (over 850) and these had grown in an ad hoc manner. To ensure the new help system achieved its expected business benefits, the client needed to structure and organise the content before it was integrated into the new interface. However, even the most dedicated user won't be happy sorting 850 cards of content, so we first had to do something to make the task manageable.

We began with a content inventory of the on-line help system. This was an important first step in describing the relationships between the different pages since it allowed us to answer questions like 'Which help pages are most commonly accessed?', 'What search terms are most common?' and 'How many help pages does the typical user view in a session?' Answers to these questions helped us classify the content into 'critical' and 'secondary' content. We also weeded out the 'ROT': content that was Redundant, Outdated or Trivial. These steps helped us reduce the sheer amount of content to something that was a bit more manageable.

Our next step was to examine the content and see if there were any obvious, stand-out topics or groups. At this point, we did in fact subject a couple of people (I was one) to the entire inventory sort to see if we could spot any obvious categories. With this approach we were able to find clusters of cards that we thought most people would place together. For example, imagine a corporate intranet that has dozens of HR policies (travel policy, environment policy, maternity policy etc). It's self-evident that most people will place these policies in the same group, so there is little to be gained by asking people to sort every policy when instead you can use a small handful of exemplars of each group in the card sort.

These two techniques helped us reduce the number of items to around 100, an acceptable number for a card sort.

As a result of our work, the new information architecture reduced the number of support enquiries from users who were unable to find or understand content. Users were now able to solve issues themselves, which indirectly increased the number of listings, sales and registrations.

### 22.13.2 Presenting the data from a closed card sort

Last year, I worked with the intranet design team in the Royal Bank of Scotland. The bank has over 150,000 employees and the design team had embarked on a major overhaul of the intranet, which contained around half a million pages. The design team wanted to check if staff could find important content in the new structure, which had close to 1000 nodes.

We carried out a closed card sort much along the lines that William describes in his article. However, we wanted to make sure that we canvassed opinions from employees in several countries, including the US, the UK and India. Because of this, we decided to use a remote, unmoderated closed card sort. We asked a representative sample of bank employees to visit a web site that contained the intranet's top-level navigation terms arranged in a tree structure (this helped us focus on navigation without the distractions of aesthetics). The participants' task was to choose the right link for various tasks, such as "Find an expenses claim form". Over 200 participants took part in the study.

The challenge with a study like this is presenting the results back to the design team in such a way that they can make an informed decision on the data. There are some obvious statistics to use — such as the number of participants who succeeded in the task — but equally useful for design is an understanding of the incorrect paths chosen by participants.

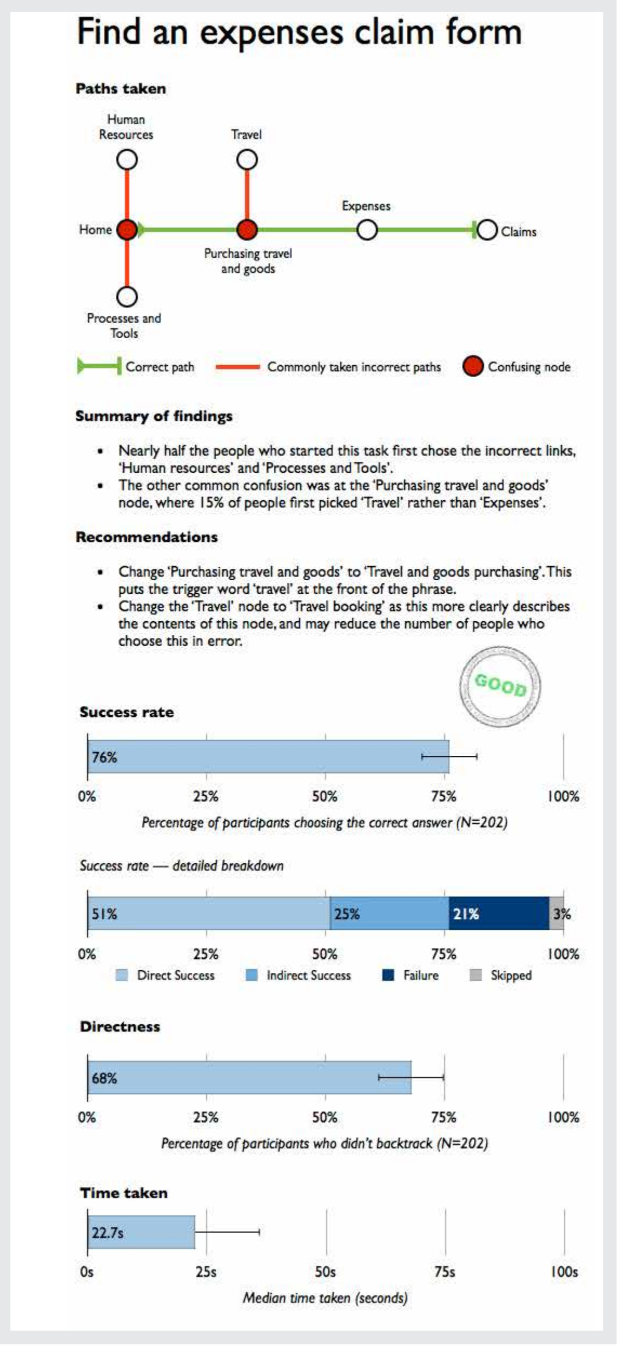



FIGURE 22.1: An example (for one task) of the way we chose to present the results.

Figure 1 shows an example (for one task) of the way we chose to present the results. Note the following features of the graphic:

- The 'tube map' diagram shows the main paths participants took to find the answer. The green line shows the correct path and the red lines show commonly taken incorrect paths. A red circle indicates a node where people chose the wrong path.
- 'Success rate' shows the percentage of participants who found the correct answer. The error bars show the 95% confidence interval.
- 'Success rate — detailed breakdown' provides more background on the success rate measure, showing how many participants needed to backtrack to find the answer ("indirect success").
- 'Directness' is the percentage of participants who didn't backtrack up the tree at any point during the task. The higher this score is, the more confident we can be that participants were sure of their answers (even if the answer is wrong). The error bars show the 95% confidence interval.
- 'Time taken' shows the median time taken by participants. The error bar shows the upper quartile. You can think of time taken as a measure of hesitation when completing the task.
- We also included a qualitative judgement on how the design performed on this task based on the measured success rate ("Very poor" through to "Excellent") and a section that interprets the findings and provides suggestions for improvement.

Other than the tube map visualisation, we were able to extract most of these metrics from the online tool we used to collect the data (Treejack). This made the analysis and presentation relatively straightforward. (Many thanks to Rebecca Shipp, RBS Group, for permission to describe this case study).

## 22.14 COMMENTARY BY CHRIS ROURKE

How to cite this commentary in your report

### Chris Rourke

© *Chris Rourke*

Chris Rourke is Managing Director & Founder of UserVision.co.uk. Chris has over 20 years commercial experience in usability, accessibility, human factors consultancy and training. He has worked with a range of clients including Hewlett Packard, Dell Computers, NCR, Houses of Parliament, Emirates Airline, DirectGov and many other commercial and public sector clients in the UK and abroad. ...

Chris Rourke
Chris Rourke is a member of The Interaction Design Foundation

One of the cruel ironies of the web is that the more information there is on your website, the harder it is to find any one single piece of information. There is more haystack to sort through to find your needle. Well, that trend is not always true, and you can at least do your best to fight that tendency by doing a very good job of organising it all. Putting things into neat, well labelled groups, and using nested hierarchies will add sense to an otherwise overwhelming mass of information.

In the UX designer's toolbox, Card Sorting is the sharpest tool for creating a sensible hierarchical structure for your information. Its cousin Tree Testing is the best for checking the robustness of that structure. Used together they are essential tools for creating a usable information architecture that is the best possible organisation to let people find their information.

William Hudson has earned a reputation as a leading thinker and practitioner in the field of card sorting, and his SynCaps software has proven very useful (and time saving) for capturing and analysing card sorting results for me and many others in the UX field.

William's Card Sorting chapter is comprehensive and educational, supported by several helpful images and a simple context that all readers will understand - the world of fruit and veg. With that as the domain, he proceeds to clearly explain

- The need for card sorting
- The types of dilemmas card sorting planners and participants encounter (e.g. the same fruit called 2 different things)
- The process for performing card sorting
- Ways to analyse the data

It is the most comprehensible and readable explanation of card sorting I have read, and will be a key learning source (along with Donna Spencer's publication which was also referenced).

In particular it provides excellent visuals to explain the outputs from card sorting. Thankfully it goes beyond presenting the tree diagram (dendrogram) which unfortunately some practitioners are tempted to take, turn 90 degrees, and exclaim: TaDah! There's my new site map, I'm all done!

More experienced practitioners will know there is a lot more that needs to be done to interpret the tree diagram, and I was especially grateful that he clearly explained that the tree diagram alone does not always tell the clearest story. For instance an item that could have strong affinities to two distinct groups could end up having an apparently moderate weak relationship to them, if the tested people were split down the middle on which they associated it with. It is a clear case where good old fashioned qualitative information from talking to people is needed to make the best decision.

In my experience, how to moderate the sessions is important and can impact the results. For instance one tip I often employ (which is helpful in the situation described above where a card has two or more natural homes) is to ask the participant to place the printed card for the item where they feel it belongs most, but if they feel that it could very comfortably fit into other groups, they can take a blank card, write that item name on it, and place it in other groups they expect it could be. All copies of the card would be processed during data capture, with SynCaps V2 splitting the item between the selected groups. As William mentions, dendrograms only support a single location for each item but the split will be apparent in the items by items and items by groups charts. The split results can be considered by the practitioner in the development of the Information Architecture, perhaps as decent locations for cross links (such as ‘see more’ type links that take the visitor to related content in other sections).

Another moderating point to consider is the amount of verbal feedback the person is to provide during the session. The core UX method of usability testing relies on a verbal stream of consciousness from the participant as they go through their journey on a website. Personally, I feel that is not appropriate for card sorting, although I recognise verbal feedback is important, especially for understanding the category names and what items are easy or difficult to sort. I usually recommend they spread out the cards to get a bird’s eye view of what they are to sort, then not disturb the participant as they see the patterns and “get in the zone”, creating their own strategy for solving this particular Information Architecture conundrum. Only after they have sorted about half the cards and applied a few labels do I try to intervene with a gentle “how’s it going?” type probe. The moderator should have that 6th sense that a hairdresser ideally has to be able to tell if the participant feels like talking or not, and not to force them to if they don’t. Once all the cards are sorted (perhaps some in a “don’t know” pile) then by all means a comprehensive debrief should be encouraged.

William gave some explanation to what I find the hardest part of card sorting – choosing what to sort when your information domain is a website with hundreds of items. Inevitably some consolidating of the items is needed, selecting only 1 or 2 representative items from what is an obviously clear group of wider items. This in itself often ends up being controversial or distracting, and always carries the risk of being used as a reason to play down the results of the card sorting (Oh yes, but you didn’t include these 3 items in the sort, it could have been very different if you had...).

Finally, my preference is always try to apply the top-down method of tree testing (or reverse car sorting, or category testing) to balance the bottom-up method of card sorting. I find tree testing to be at least as useful a method to get clients to see the importance of a good, user-centred information architecture. After all, the process of tree testing is far more similar to the way people actually forage for information while navigating on a website. Furthermore the quantitative and statistical data that comes from it is very compelling especially when it can be done before and after a revision to the Information Architecture (“previously this topic was found by 50% of people without any errors, now that is up to 75%”). It can also be done remotely, and other resources in addition to the ones mentioned in the chapter include NaviewApp and UserZoom. Williams core area is card sorting, but if more could be presented on the top-down method perhaps it could be re-titled Card Sorting and Information Architecture research.

William Hudson's chapter is nonetheless comprehensive and meets the need of those new to card sorting and those with some experience. It will definitely be a valuable reference to those looking to implement this research to improve their site navigation and Information Architecture.

## 22.15 REFERENCES

*Bergström, John A. (0000b): An Experimental Study of Some of the Conditions of Mental Activity. In The American Journal of Psychology, 6 (2) pp. 247-274*

*Bergström, John A. (0000a): Experiments upon Physiological Memory by Means of the Interference of Associations. In The American Journal of Psychology, pp. 356-369*

*Bernard, H. Russell and Ryan, Gery W. (2009): Analyzing Qualitative Data: Systematic Approaches. Sage Publications, Inc*

*Capra, Miranda G. (2005): Factor analysis of card sort data: an alternative to hierarchical cluster analysis. In Human Factors, 49 (5) pp. 691-695*

*Corter, James E. (1996): Tree Models of Similarity and Association (Quantitative Applications in the Social Sciences). Sage Publications, Inc*

*Coxon, Anthony P. M. (1999): Sorting Data: Collection and Analysis (Quantitative Applications in the Social Sciences). Sage Publications, Inc*

*Dong, Jianming, Martin, Shirley and Waldo, Paul (2001): A user input and analysis tool for information architecture. In CHI 01 extended abstracts on Human factors in computing systems CHI 01,*

*Dubois, Cornelius (1949): The Card-Sorting or Psychophysical Interview. In Public Opinion Quarterly, 13 (4) pp. 619-628*

*Eling, Paul, Derckx, Kristianne and Maes, Roald (2008): On the historical and conceptual background of the Wisconsin Card Sorting Test. In Brain and Cognition, 67 (3) pp. 247-253*

*Fincher, Sally and Tenenberg, Josh (2005): Making sense of card sorting data. In Expert Systems, 22 (3) pp. 89-93*

*Giovannini, Peter (2012). How to carry out pile sorting and how to analyse the data with Anthropac: a tutorial. Retrieved 2 January 2012 from http://petergiovannini.com/ethnobotany-methods/how-to-pile-sorting-with-anthropac-tutorial.html*

*Harloff, Joachim (2005): Multiple Level Weighted Card Sorting. In Methodology: European Journal of Research Methods for the Behavioral and Social Sciences, 1 (4) pp. 119-128*

*Jastrow, Joseph (0000): A sorting apparatus for the study of reaction-times. In Psychological Review, 5 (3) pp. 279-285*

*Kiger, John I. (1984): The Depth/Breadth Trade-Off in the Design of Menu-Driven User Interfaces. In International Journal of Man-Machine Studies, 20 (2) pp. 201-213*

*Klare, George R. (1963): Measurement of Readability. Umi Research Pr*

*Kline, Linus W. and Kellogg, Chester E. (1914): Cards as Psychological Apparatus. In Science, 39 pp. 657-659*

*Miller, George A. (1969): A psychological method to investigate verbal concepts. In Journal of Mathematical Psychology, 6 (2) pp. 169-191*

*Morville, Peter and Rosenfeld, Louis (2006): Information Architecture for the World Wide Web: Designing Large-Scale Web Sites. OReilly Media*

*Nielsen, Jakob (2004): Card Sorting : How Many Users to Test. In Useit Alertbox, (0) pp. 3-7*

*Nielsen, Jakob and Sano, Darrell (1995): SunWeb: user interface design for Sun Microsystem's internal Web. In Computer Networks and ISDN Systems, 28 (1) pp. 179-188*

*Romesburg, Charles (2004): Cluster Analysis for Researchers. Lulu.com*

*Sachs, Joe (2002): Aristotle's Metaphysics. Green Lion Press*

*Spencer, Donna (2009): Card Sorting. Rosenfeld Media*

*Tullis, Thomas S. (1985): Designing a Menu-Based Interface to an Operating System. In: Borman, Lorraine and Curtis, Bill (eds.) Proceedings of the ACM CHI 85 Human Factors in Computing Systems Conference April 14-18, 1985, San Francisco, California. pp. 79-84*

*Tullis, Tom and Wood, Larry E. (2004): How Many Users Are Enough for a Card-Sorting Study? The Card-sorting Study. In Learning, pp. 1-10*

*Weller, Susan C. and Romney, A. Kimball (1988): Systematic Data Collection (Qualitative Research Methods Series 10). Sage Publications, Inc*

*Wood, Jed R. and Wood, Larry E. (2008): Card Sorting: Current Practices and Beyond. In Journal of Usability Studies, 4 (1) pp. 1-6*

## About the author

**William Hudson**

Personal Homepage:
http://www.syntagm.co.uk/

Current place of employment:
Syntagm

William Hudson is a User Experience Strategist who consults, writes and teaches in the fields of user-centred design, user experience and usability. He has over 40 years experience in the development of interactive systems, initially with a background in software engineering. William was the product and user interface designer for the Emmy-award-winning “boujou”; now an indispensible tool in many film studios. He has specialized in interaction design and human-computer interaction since the late 1980’s. William has written and taught courses which have been presented to hundreds of software and web developers, designers and managers in the UK, North America and Europe. He has developed and presented courses for the Nielsen Norman Group. William is the founder and principal consultant of Syntagm, a consultancy specializing in the design of interactive systems established in 1985.

## Your notes and thoughts on chapter 22

Record your notes and thoughts on this chapter. If you want to share these thoughts with others online, go to the bottom of the page at:
http://www.interaction-design.org/encyclopedia/card_sorting.html

### NOTES

# CHAPTER 23

# Wearable Computing

*by Steve Mann.*

Wearable computing is the study or practice of inventing, designing, building, or using miniature body-borne computational and sensory devices. Wearable computers may be worn under, over, or in clothing, or may also be themselves clothes (i.e. "Smart Clothing" (Mann, 1996a))

## 23.1 BEARABLE COMPUTING

The field of wearable computing, however, extends beyond "Smart Clothing". The author often uses the term "Body-Borne Computing" or "Bearable Computing" as a substitute for "Wearable Computing" so as to include all manner of technology that is on or in the body, e.g. implantable devices as well as portable devices like smartphones. In fact the word "portable" comes from the French word "porter" which means "to wear".

## 23.2 PRACTICAL APPLICATIONS

Applications of body-borne computing include seeing aids for the blind or visually impaired, as well as memory aids to help persons with special needs. The MindMesh, an EEG (ElectroEncephaloGram) based “thinking cap”, for example, allows the user to plug various devices into their brain. A blind person can plug in a camera and use it as an “eye”.

Moreover, body-borne computing in the inclusive sense is for everyone, in the form of such applications as wayfinding, and Personal Safety Devices (PSDs). Body-borne computing is already a part of many people’s lives, in the form of a smartphone that helps them find their way if they get lost, or helps protect them from danger (e.g. for emergency notification). The next generation of smartphones will be borne by the body in a way that it is always attentive (e.g. that the camera can always “see” one’s environment), so that if a person gets lost, the device will help the user “remember” where they are. Additionally, it will function like the “black box” flight recorder on an aircraft, and, in the event of danger, will be able to automatically notify others of the user’s physiological state as well as what happened in the environment.

Consider, for example, a simple heart monitor that continuously records ECG (ElectroCardioGram) along with video of the environment. This may help physicians correlate heart arrythmia, or other irregularties, with possible envioronmental causes of stress - a physician may be able to see what was happening to the patient at the time a problem was first detected.

## 23.3 WEARABLE COMPUTING AS A RECIPROCAL RELATIONSHIP BETWEEN MAN AND MACHINE

An important distinction between wearable computers and portable computers (handheld and laptop computers for example) is that the goal of wearable computing is to position or contextualize the computer in such a way that the human and computer are inextricably intertwined, so as to achieve Humanistic Intelligence – i.e. intelligence that arises by having the human being in the feedback loop of the computational process, e.g. Mann 1998.

An example of Humanistic Intelligence is the wearable face recognizer (Mann 1996) in which the computer takes the form of electric eyeglasses that "see" everything the wearer sees, and therefore the computer can interact serendipitously. A handheld or laptop computer would not provide the same serendipitous or unexpected interaction, whereas the wearable computer can pop-up virtual nametags if it ever "sees" someone its owner knows or ought to know.

In this sense, wearable computing can be defined as an embodiment of, or an attempt to embody, Humanistic Intelligence. This definition also allows for the possibility of some or all of the technology to be implanted inside the body, thus broadening from "wearable computing" to "bearable computing" (i.e. body-borne computing).

One of the main features of Humanistic Intelligence is constancy of interaction, that the human and computer are inextricably intertwined. This arises from constancy of interaction between the human and computer, i.e. there is no need to turn the device on prior to engaging it (thus, serendipity).

Another feature of Humanistic Intelligence is the ability to multi-task. It is not necessary for a person to stop what they are doing to use a wearable computer because it is always running in the background, so as to augment or mediate the human's interactions. Wearable computers can be incorporated by the user to act like a prosthetic, thus forming a true extension of the user's mind and body.

It is common in the field of Human-Computer I nteraction (HCI) to think of the human and computer as separate entities. The term "Human-Computer Interaction" emphasizes this separateness by treating the human and computer as different entities that interact. However, Humanistic Intelligence theory thinks of the wearer and the computer with its associated input and output facilities not as separate entities, but regards the computer as a second brain and its sensory modalities as additional senses, in which synthetic synesthesia merges with the wearer's senses. In this context, wearable computing has been referred to as a

"Sixth Sense" (Mann and Niedzviecki 2001, Mann 2001 and Geary 2002).

When a wearable computer functions as a successful embodiment of Humanistic Intelligence, the computer uses the human's mind and body as one of its peripherals, just as the human uses the computer as a peripheral. This reciprocal relationship is at the heart of Humanistic Intelligence (Mann 2001, Mann 1998, and Knight 2000)

## 23.4 CONCRETE EXAMPLES OF WEARABLE COMPUTING

### 23.4.1 Example 1: Augmented Reality

Augmented Reality means to super-impose an extra layer on a real-world environment, thereby *augmenting* it. An "augmented reality" is thus a view of a physical, real-world environment whose elements are augmented by computer-generated sensory input such as sound, video, graphics or GPS data. One example is the Wikitude application for the iPhone which lets you point your iPhone's camera at something, which is then "*augmented*" with information from the Wikipedia (strictly speaking this is a mediated reality because the iPhone actually modifies vision in some ways - even if nothing more than the fact we're seeing with a camera).

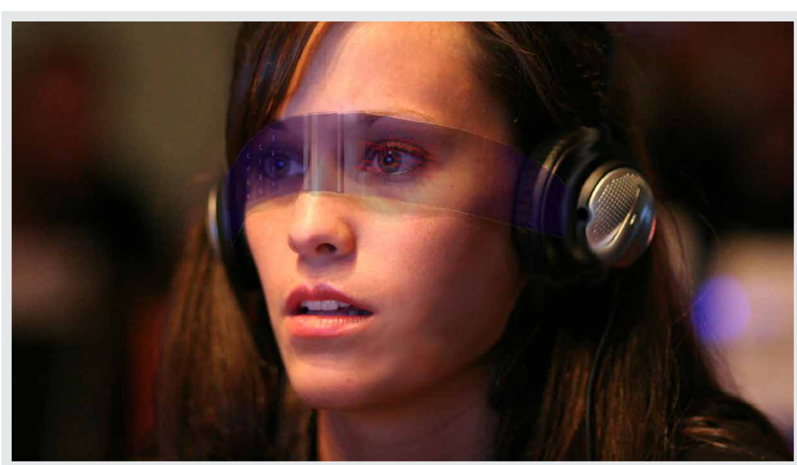


Figure 23.1: Augmented Reality prototype.

FIGURE 23.2: Photograph of the Head-Up Display taken by a pilot on a McDonnell Douglas F/A-18 Hornet.

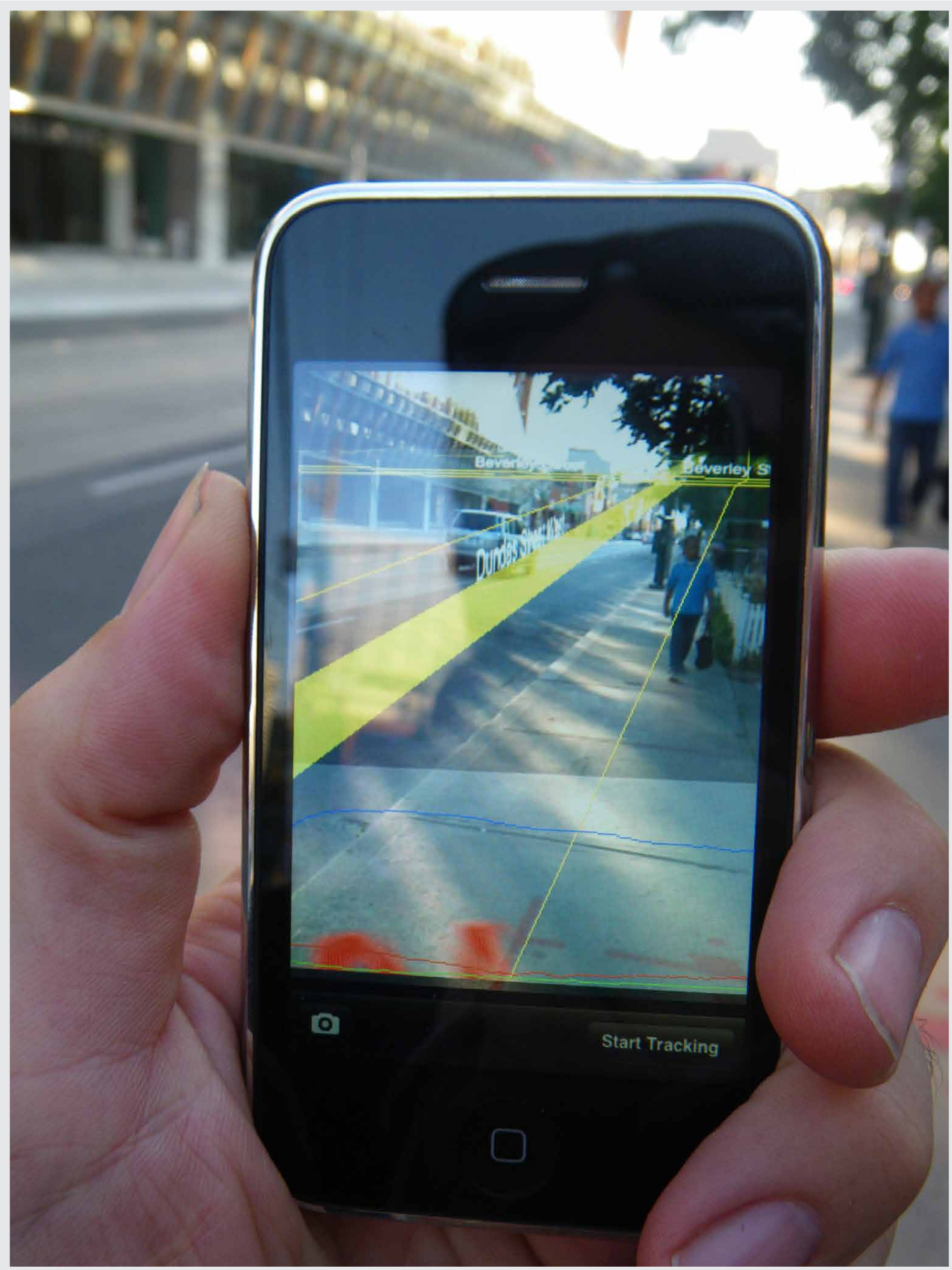



FIGURE 23.3: The glogger.mobi application: Augmented reality 'lined up' with reality.


Figure 23.4: The Wikitude iphone application.



A concrete example of wearable computing used for augmented reality is Mann's pendant-based camera and projector system for Augmented Reality. The system shown below was completed by Mann in 1998:

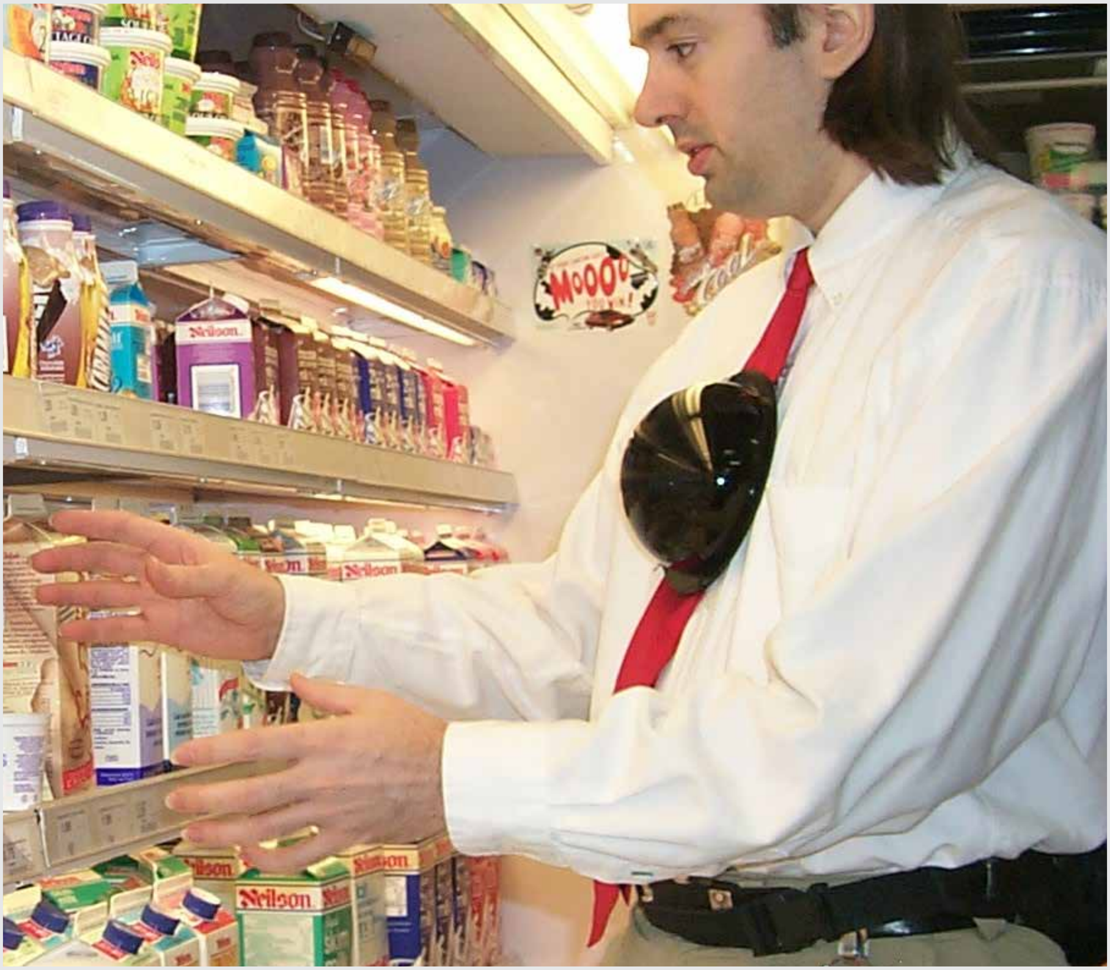



FIGURE 23.5: Neckworn self-gesturing webcam and projector system designed and built by Steve Mann in 1998.

*Courtesy of Steve Mann. Copyright: CC-Att-SA-3 (Creative Commons Attribution-ShareAlike 3.0).*

Figure 23.6: Closeup of dome pendant showing the laser-based infinite depth-of-focus projector, called an "aremac" (Mann 2001). The laser-based aremac was developed to project onto any 3D surface and does not require any focusing adjustments.

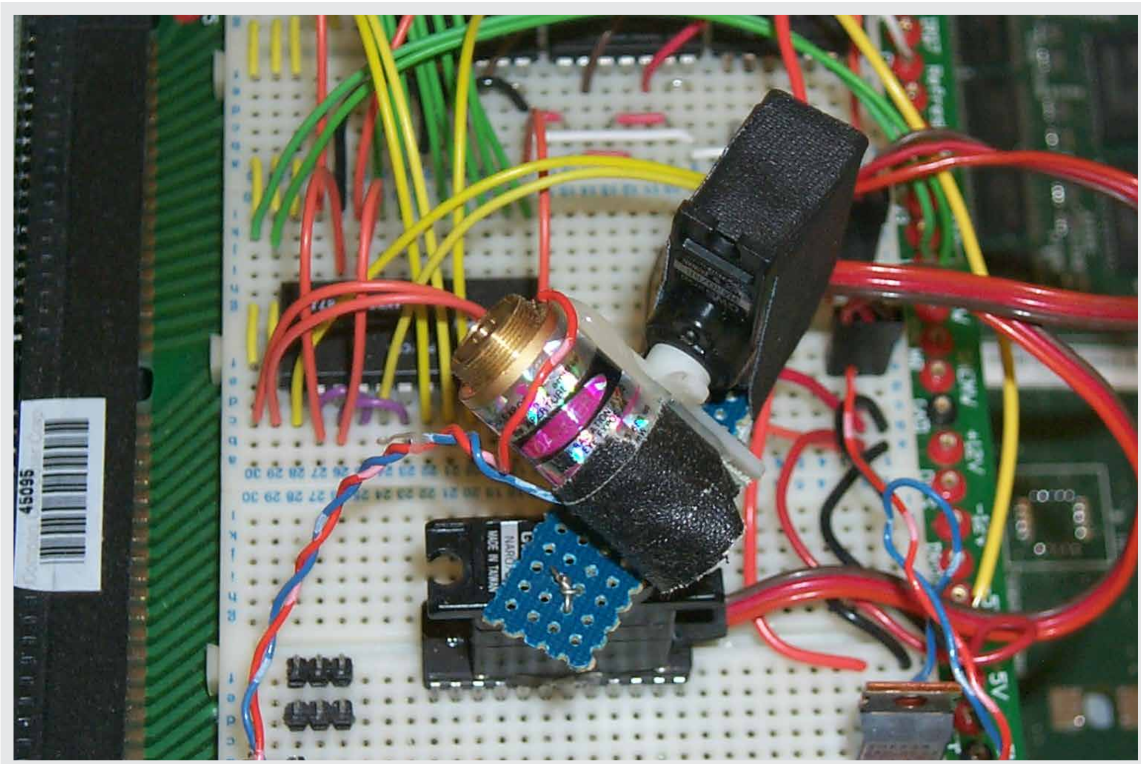


Figure 23.7: Early breadboard prototype of the aremac that Mann developed for the neckworn webcam+projector.



In Figure 23.5 the wearer is shopping for milk, but this could also have been a more significant purchase like a new car or a house. The wearer's wife, at a remote location, is looking through the camera by way of a projection screen in her living room in another country. She points a laser pointer at the screen, and a vision system in the projector tracks that and remotely operates the aremac in the wearer's necklace. Thus he sees whatever she draws or scribbles on her screen. This scribbling or drawing directly annotates the "reality" that he's experiencing.

In another application, the wearer can use hand gestures to control the wearable computer. The author referred to this system as "Synthetic Synesthesia of the Sixth Sense", and it is often called "SixthSense" for short.

This wearable computer system was used as a teaching example at University of Toronto, where hundreds of students were taught how to build the system, including the vector-graphics laser-based infinite depth-of-field projector, using surplus components obtained at low cost. The system cost approximately $75 for each student to build (not including the computer). The software and circuit board design for this system was distributed to students under an Open Source licence, and the circuit board itself was designed using Open Source computer programs (PCB, kicad, etc.), see Mann 2001b.

### 23.4.2 Example 2: Diminished Reality

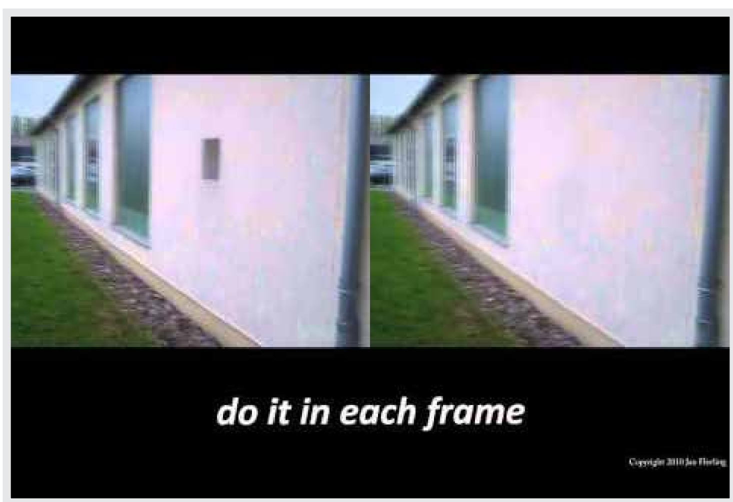
While the goal of Augmented Reality is to *augment* reality, an Augmented Reality system often accomplishes quite the opposite. For example, Augmented Reality often adds to the confusion of an already confusing existence, adding extra clutter to an already cluttered world. There seems to be a fine line between Augmented Reality and information overload.

Sometimes there are situations where it is appropriate to remove or diminish clutter. For example, the electric eyeglasses (www.eyetap.org) can assist the visually impaired by simplifying rather than complexifying visual input. To do this, visual reality can be re-drawn as a high-contrast cartoon-like world where lines and edges are made more bold and crisp and clear, thus being visible to a person with limited vision.

Another situation in which diminished reality makes sense is dealing with advertising. Our world is increasingly being cluttered with advertising and visual detritus. The electric eyeglasses can filter out unwanted advertising, and reclaim that visual space for useful information. Unwanted advertising, seen once, is inserted into a killfile (e.g. a file of particular ads that are to be reclaimed). For example, if the user is a non-smoker, he or she may decide to put certain cigarette ads into the killfile, so that when subsequently seen, they are removed. That space can then be overwritten with useful data. The following videos show examples:


Video 23.1 Diminished Reality concept video by Steve Mann and James Fung from 2008. Implementation was done on an eyetap device.

*Courtesy of Steve Mann. Copyright: CC-Att-ND (Creative Commons Attribution-NoDerivs 3.0 Unported). View full screen or download (0)*




Video 23.2 Diminished Reality video by Jan Herling and Wolfgang Broll 2010.

*Courtesy of US Senate. Copyright: pd (Public Domain (information that is common property and contains no original authorship)).*

*Courtesy of US Senate (derivative work). Copyright: pd (Public Domain (information that is common property and contains no original authorship)).*

FIGURE 23.8 A-B: Simplifying rather than complexifying visual input. Such "diminished reality" may help the visually impaired.

### 23.4.3 Example 3: Mediated Reality

Another concrete example is Mediated Reality. Whereas the Augmented Reality system shown above can only add to “reality”, the Mediated Reality systems can augment, deliberately diminish, or otherwise enhance or modify visual reality beyond what is possible with Augmented Reality. Thus mediated reality is a proper superset of augmented reality.

Mediated Reality refers to a general framework for artificial modification of human perception by way of devices for augmenting, deliberately diminishing, and more generally, for otherwise altering sensory input. A simple example is electric eyeglasses (www.eyetap.org) in which the eyeglass prescription is downloaded wirelessly, and can be updated continuously in a way that’s subject-matter specific or task-specific.

These electric eyeglasses also allow the wearers to reconfigure their vision into different spectral bands. For example, infrared eyeglasses allow us to see where people have recently stood on the ground (where the ground is still warm) or which cars in a parking lot recently arrived (because the engine is still warm). One can see how well the insulation in a building is doing, by observing where heat is leaking out of the building. A roofer can see where a roof membrane may be problematic, or where heat is leaking out of a building. Moreover, during roof repair, one can see the molten asphalt, and get a good sense of whether or not it is at the right temperature.

The electric eyeglasses can allow us to see in different spectral bands while actually repairing a roof, thus forming a closed feedback look, as an example of Humanistic Intelligence. See Figure 23.8 A-B.

*Courtesy of Steve Mann. Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.*

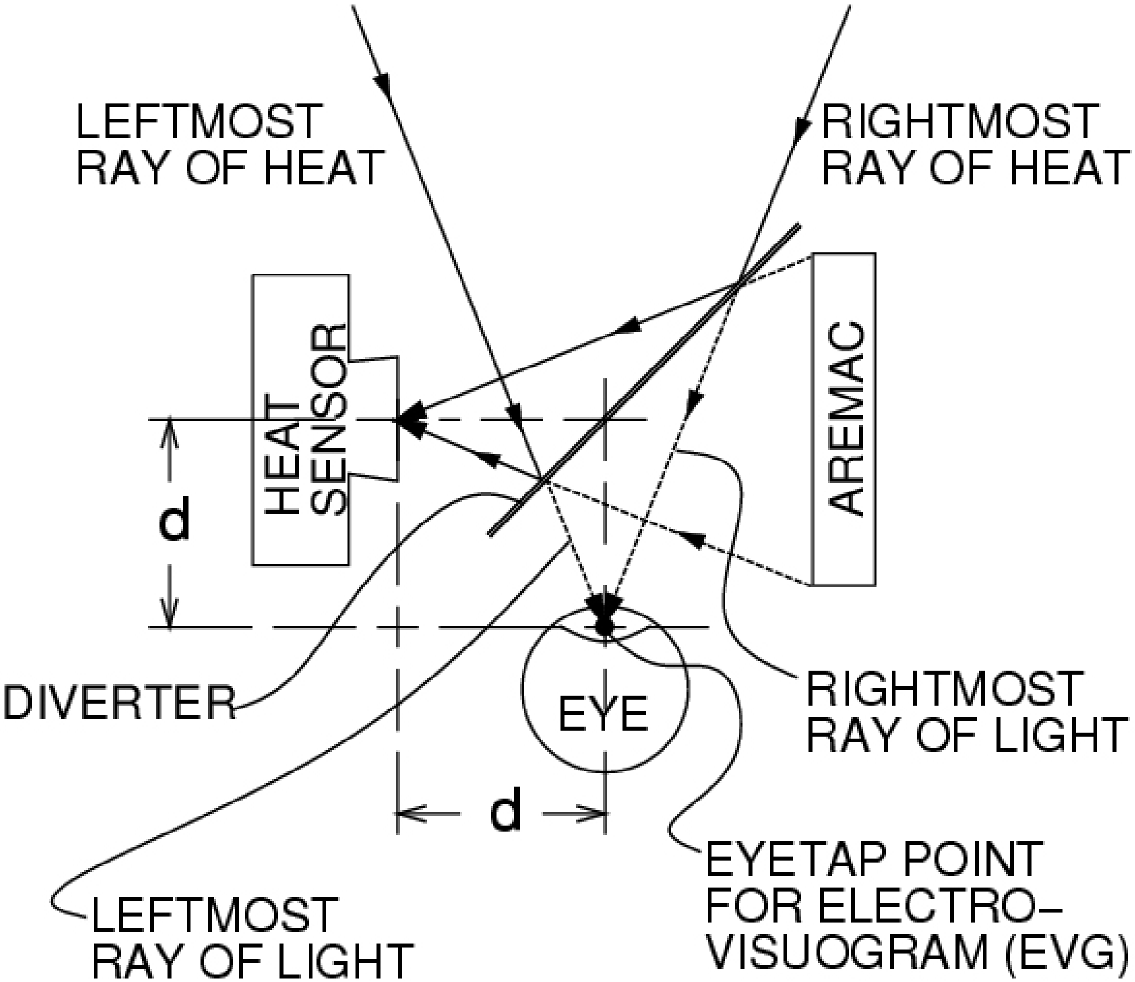





Figure 23.9 A-B: A (at left): Author (looking down at the mop he is holding) wearing a thermal EyeTap wearable computer system for seeing heat. This device modified the author's visual perception of the world, and also allowed others to communicate with the author by modifying his visual perception. A bucket of 500 degree asphalt is present in the foreground. B (at right): Thermal EyeTap principle of operation: Rays of thermal energy that would otherwise pass through the center of projection of the eye (EYE) are diverted by a specially made 45 degree "hot mirror" (DIVERTER) that reflects heat, into a heat sensor. This effectively locates the heat sensor at the center of projection of the eye (EYETAP POINT). A computer controlled light synthesizer (AREMAC) is controlled by a wearable computer to reconstruct rays of heat as rays of visible light that are each collinear with the corresponding ray of heat. The principal point on the diverter is equidistant to the center of the iris of the eye and the center

of projection of the sensor (HEAT SENSOR). (This distance, denoted “d”, is called the eyetap distance.) The light synthesizer (AREMAC) is also used to draw on the wearer’s retina, under computer program control, to facilitate communication with (including annotation by) a remote roofing expert.

## 23.5 HISTORY OF WEARABLE COMPUTING

Depending on how broadly wearable computing is defined, the first wearable computer might have been an abacus hung around the neck on a string for convenience, or worn on the finger.

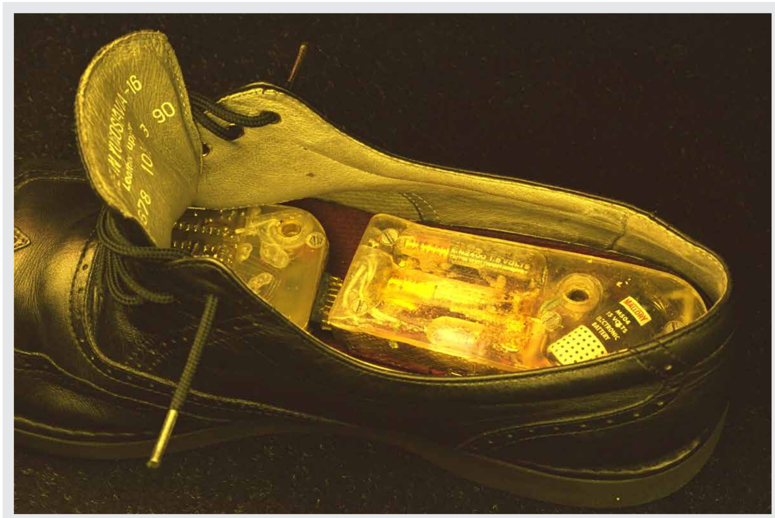
Or it might have been the pocket watches of the early 1500s, or the wristwatches that replaced them, since a timepiece is a computer of sorts (i.e. a device that computes or keeps time). See Figure 23.10 A-B.



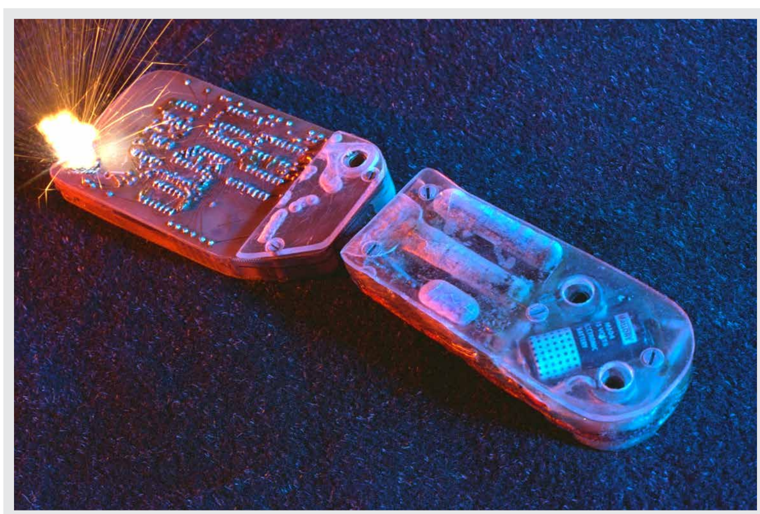
Figure 23.10 A-B: Leftmost, One of the first pocket watches, called "The Nuremberg Egg", made arround 1510. Rightmost, an early digital wristwatch from the 1920s.

More recently electronic calculators (which could be carried in a pocket or worn on the wrist) emerged, as did electronic timepieces. Other task-specific electronic circuits included a timing device concealed in a shoe to help the wearer cheat at a game of roulette (Bass 1985).

A common understanding of the term "computer" is that a computer is something that is programmable by the user, while it is being used, or that is of a relatively general-purpose nature (e.g. the user can change programs and run various applications).



*Courtesy of Steve Mann. Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.*



*Courtesy of Steve Mann. Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.*

Figure 23.11 A-B: A timing device designed to be concealed in a shoe for use in roulette invented by Ed Thorp and Claude Shannon in 1961 but first mentioned in Thorp 1966. Although it uses electronic circuits it could not be programmed by the wearer, and ran only one application: a program that computed time. The devices described above are predecessors to what is commonly meant by the term "wearable computer".

Figure 23.12: Here is a "computer" (an abacus) and since it is a piece of jewelry (a ring), it is wearable. Such devices have existed for centuries, but do not successfully embody Humanistic Intelligence. In particular, because the abacus is task-specific, it does not give rise to what we generally mean by "wearable computer". For example, its functions and purpose (algorithms, applications, etc.) can't be reconfigured (programmed) by the end user while wearing it. In short, "wearable computer" means more than the sum of its parts i.e. more than just "wearable" and "computer". Made with beads of a silver ring abacus of 1.2 centimeter long and 0.7 centimeter wide, dating back to Chinese Qing Dynasty (1616-1911 BC).



Thus a task-specific device like an abacus or wristwatch or timer hidden in a shoe is not generally what we think of when we think of "computer". Indeed, what made the computer revolution so profound was that the computer is a software re-programmable device capable of being used for a wide variety of complex algorithms and applications.

In the 1970s and early 1980s Steve Mann designed and built a number of general-purpose wearable computer systems, including various kinds of sensing, biofeedback, and multimedia computers such as wearable musical instruments, audio-based computers, and seeing aids for the blind.

In 1981 Mann designed and built a backpack-based general-purpose multimedia wearable computer system with a head-mounted display visible to one eye. The system provided text, graphics, audio, and video capability, and included a hand-held chording keyer (for one-handed input). Because of its generality, this system fit the description of what most people would call a "computer" by today's standards.

The system allowed various computer applications to be run while walking around doing other things. The computer could even be programmed (i.e. new applications could be written) while walking around. Among the applications written for this wearable computer system was an application for photographically mediated reality and "lightvector painting" ("lightvectoring") used extensively throughout the 1980s. A variety of different systems were designed and built by Mann in the 1980s, and this marked a steady evolution in wearable computing toward something resembling ordinary eyeglasses by the late 1990s (Mann 2001b).

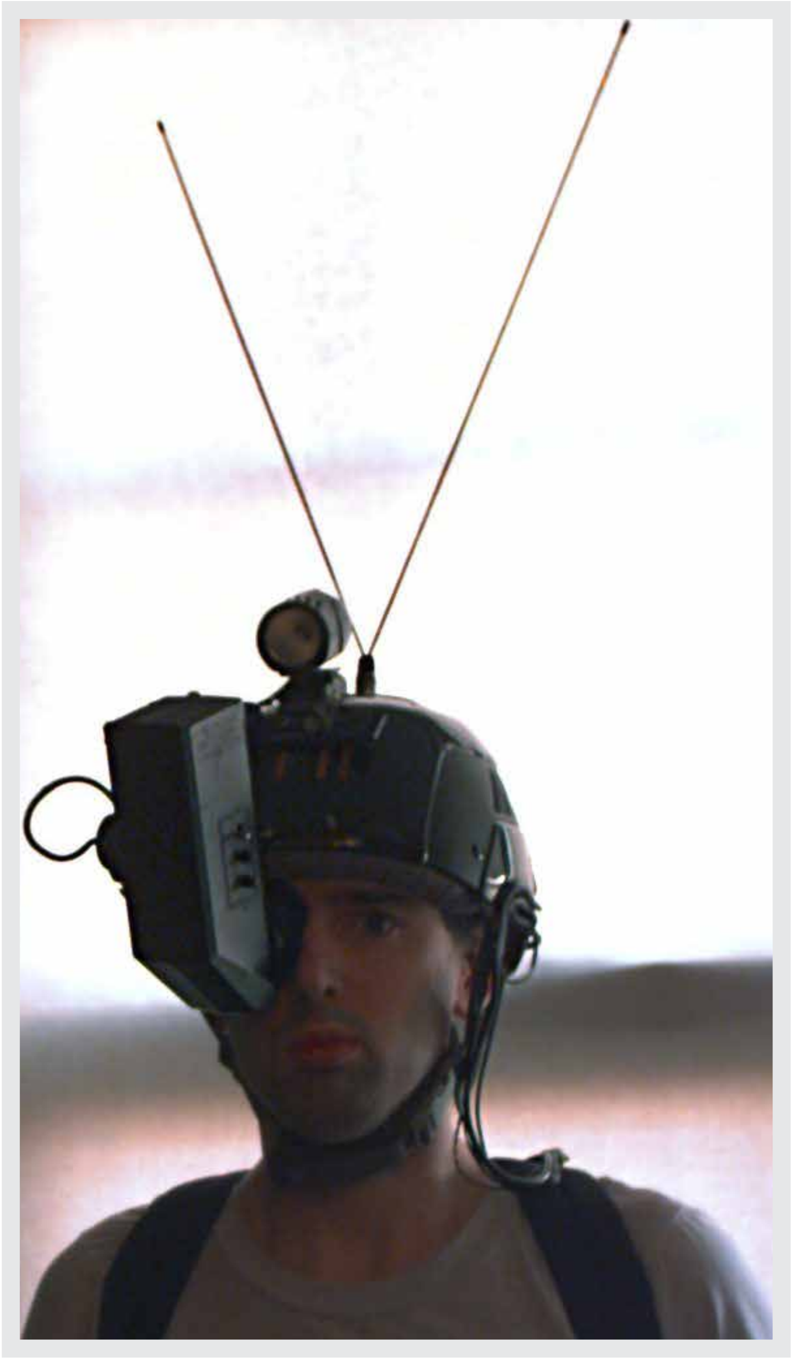




Figure 23.13 A-B: The WearComp wearable computer by the late 1970s and early 1980s - a backpack based system.



Figure 23.14 A-B: The WearComp wearable computer by the mid 1980s.


M.I.T.
MASSACHUSETTS



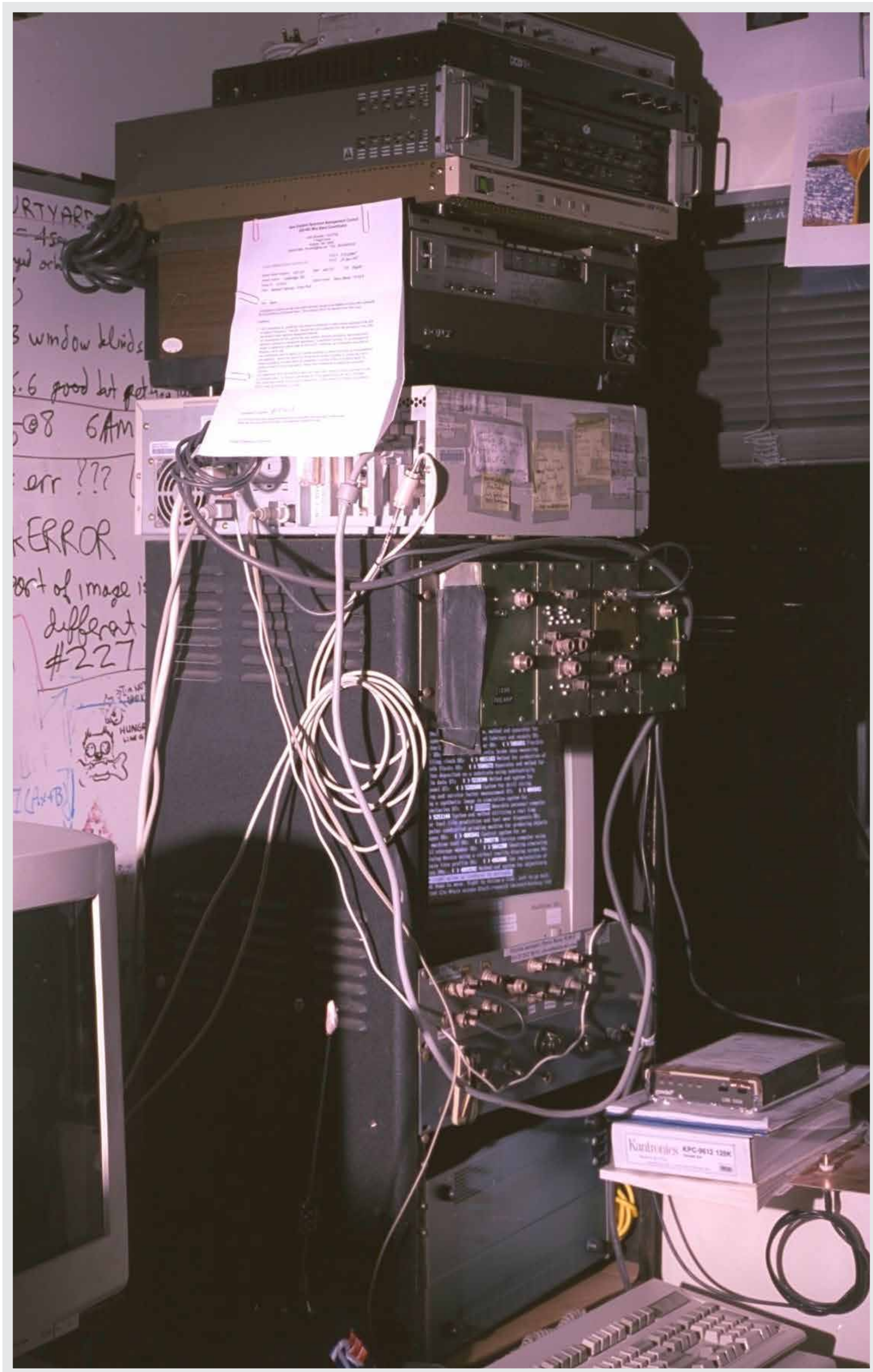

FIGURE 23.15 A-B: The WearComp wearable computer anno 1990 (leftmost) and by the mid 1990s (rightmost).




FIGURE 23.16 A-B: The WearComp wearable computer by the late 1990s, resembling ordinary eyeglasses.

By 1994 Mann had streaming live video from his wearable computer to and from the World Wide Web, such that viewers to his web site could see what he was seeing, as well as annotate what he was seeing (i.e. "scribble on his retina" so to speak). This "Wearable Wireless Webcam" was the first embodiment of live webcasting from a wireless device.

Because there were no wireless service providers at this time (much of this technology had not been invented yet), it all had to be built by hand. See Figure 23.17 A-B-C.


6AM
ERROR
different
#227
Kantronics

*Courtesy of Steve Mann. Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.*

*Courtesy of Steve Mann. Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.*

FIGURE 23.17 A-B-C: Early 1990s wireless communications system invented, designed, and built by Mann. Home-made wireless network (left) 19-inch relay rack with various equipment including microwave link to+from the roof of the tallest building in the city (middle) Steve Mann wearing the computer system (electric eyeglasses) while servicing the antenna on the roof of the tallest building in the city. This along with a network of other antennas, was setup to obtain wireless connectivity. Mann applied for and obtained a 100kHz spectral allocation through the New England Spectrum Management Council, 445.225MHz, for a community of "cyborgs". In many ways amateur radio (ham radio) was the predecessor of the modern Internet, where radio operators would actively communicate from their homes (base stations), vehicles (mobile units), or bodies (portable units) with other radio operators around the world. Mann was an active ham radio operator, with callsign N1NLF.

Another ham radio operator, Steven K. Roberts, callsign N4RVE, designed and built Winnebiko-II, a recumbent bicycle with on-board computer and chording keyer. Roberts referred to his efforts as "nomadness", which he defined as "nomadic computing". For example, he could type while riding the bicycle (Roberts 1988).

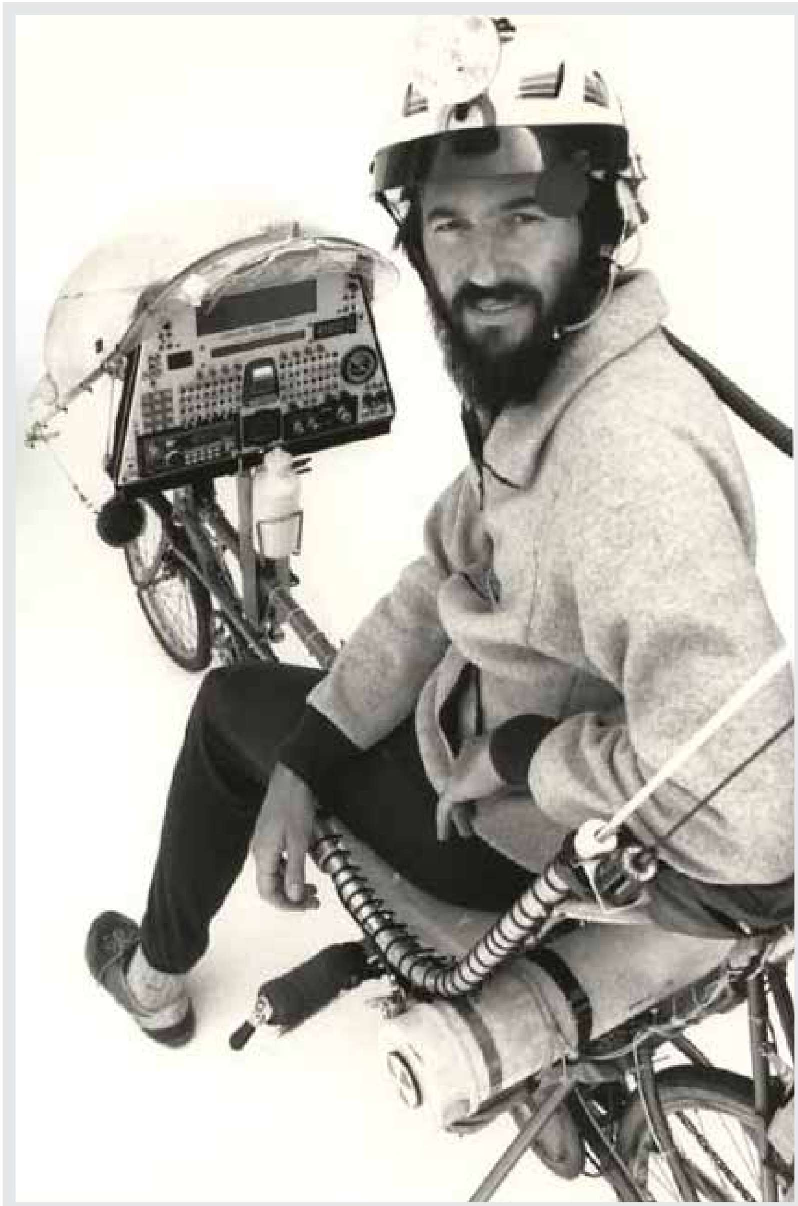


FIGURE 23.18: The Winnebiko II system, which integrated a wide range of computer and communication systems in such a way that they could be effectively be used while riding, including a chord keyboard in the handlebars.

In 1989 a "pushbroom" type display using a linear array of 280 red LEDs became available from a company called Reflection Technology. The product was referred to as the Private Eye. Because of the lack of adequate grayscale on the 280 LEDs, and also due to the use of red light (which makes image display difficult to see clearly), the Private Eye was aimed mainly at text display, rather than the multimedia computing typical of the earlier wearable computing efforts.

However, despite its limitations, the Private Eye display product brought wearable computing to the mainstream, making it easy for hobbyists to put together a wearable computer from commercial off-the-shelf devices.

Among these hobbyists were Gerald "Chip" Maguire from Columbia University and Doug Platt who built a system he called the "Hip-PC", (Bade et al 1990) and later, Thad Starner at MIT.

In 1993, Starner built a system based on Platt's design, using a Private Eye display and a Handykey Twiddler keyer. That same year, Steve Feiner, et al, at Columbia University created an augmented reality system based on the Private Eye (Feiner et al 1993).

By 1990 Xybernaut Corporation was founded, originally called Computer Products & Services Incorporated (CPSI), the name being changed to Xybernaut in 1996. Xybernaut marketed wearable computing in vertical market segments such as to telephone repair technicians, soldiers, and the like. Around this time, another company, ViA Inc., produced a flexible wearable computer that could be worn like a belt, although there were some problems with the "rubber dockey" product that connected it to the outside world.

In 1998 Steve Mann made a working prototype of a wristwatch computer running GNU Linux. The wristwatch included video-conferencing capability and was demonstrated at the ISSCC 2000 conference in February. In July 2000, Mann's Linux wristwatch was featured on the cover of Linux Journal, Issue 75, along with an article about it. See Figure 23.19.


Figure 23.19: A wristwatch computer with videoconferencing capability running the videoconferencing application underneath a transparent oclock, running XF86 under the GNUX (GNU+Linux) operating system. The computer, being general-purpose in nature, rather than task-specific (e.g. beyond merely keeping time, etc.) made this device fit what we typically mean by "wearable computer" (i.e. something that the wearer can reconfigure, program, etc., while wearing it, as well as something that implements Humanistic Intelligence). The project was completed in 1998. The SECRET function, when selected, conceals the videoconferencing window by turning off the transparency of the o'clock, so that the watch then looks like an ordinary watch (just showing the clock filling the entire 640x480 pixel screen). The OPEN function cancels the SECRET function and opens the videoconferencing session up again. The system streamed live video at 7fps, 640x480, 24 bit color.

*Courtesy of Steve Mann. Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.*

In 2001 IBM publicly unveiled a prototype for a wristwatch computer running Linux, but it has yet to be commercially released.

The vision of wearable computing has yet to be fulfilled commercially, but the proliferation of portable devices such as smart phones suggests an evolution in that direction. Most notably, with the appropriate input and output devices, a smart phone can form a good central processor upon which to realize an embodiment of Humanistic Intelligence.

## 23.6 WEARABLE COMPUTING INPUT OUTPUT DEVICES

For a wearable computer to achieve a full implementation of Humanistic Intelligence there needs to be a constancy of user interaction, or at least a low threshold for interaction to begin. Much of the serendipity is lost if the computer must be taken out of a purse or pocket and started up). Therefore a wearable computer typically has an output device such as a display that the user can sense, and an input device with which to communicate explicitly with the computer.

Starting with the input device, the first wearable computers used a keying device called a “keyer”. The keyer is inspired by the telegraph keyer of ham radio (e.g. a morse code input device), which has evolved from the single key, then to iambic (or what the author calls “biambic”), then to triambic, and more generally, multiambic. The term “iambic” existed previously to describe two-key morse code devices (e.g. morse code comprised of iambs, i.e. concepts of rhythm borrowed from poetry, having meter of verse comprised of iambs). Mann, upon hearing the word “iambic” in childhood, misunderstood the term “iambic” and thought it meant “biambic” and due to his mistake, he generalized the concept to “triambic” (3 buttons), and so on

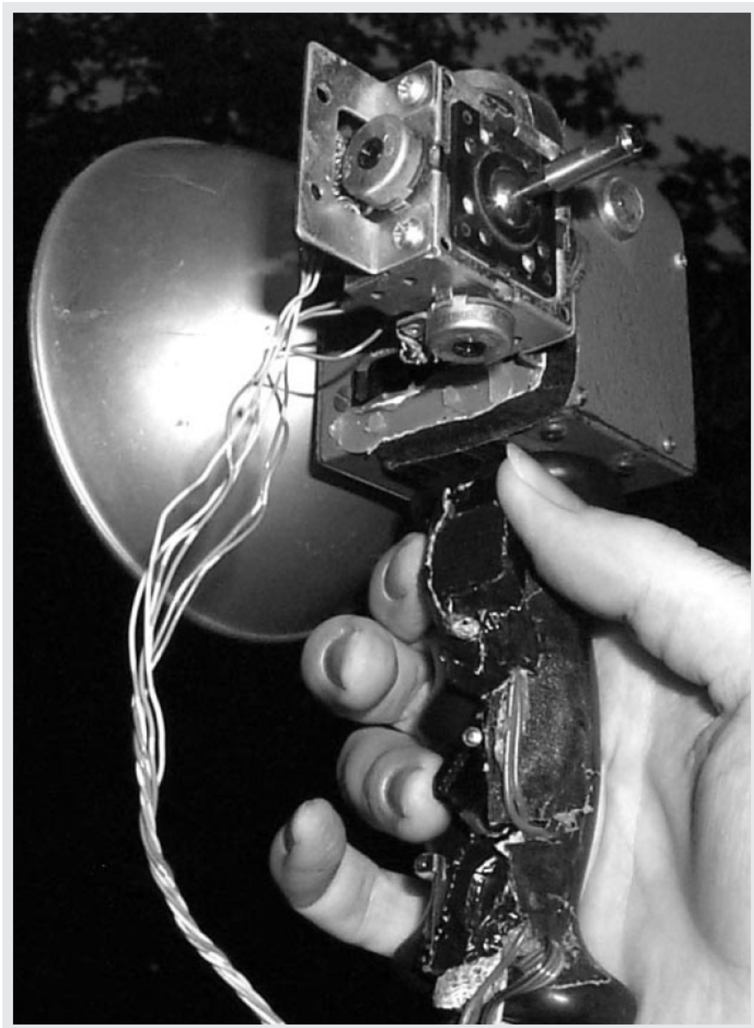




Figure 23.20 A-B: Early (1978) wearable computing keyer prototypes invented, designed, and built by the author. These keyers were built into the hand grip of another device. Leftmost, is the author's PoV (Persistence of Vision) pushbroom text generator. Text keyed into the keyer was displayed on a linear array of lights waved through the air like a pushbroom, visible in a dimly lit space, either by persistence of human vision, or by long exposure photographs. Rightmost, a "painting with lightvectors" invention allows various layers to be built in a multidimensional "lightspace". Note the keyer combined with the pointing device, which was connected to a multimedia wearable computer.

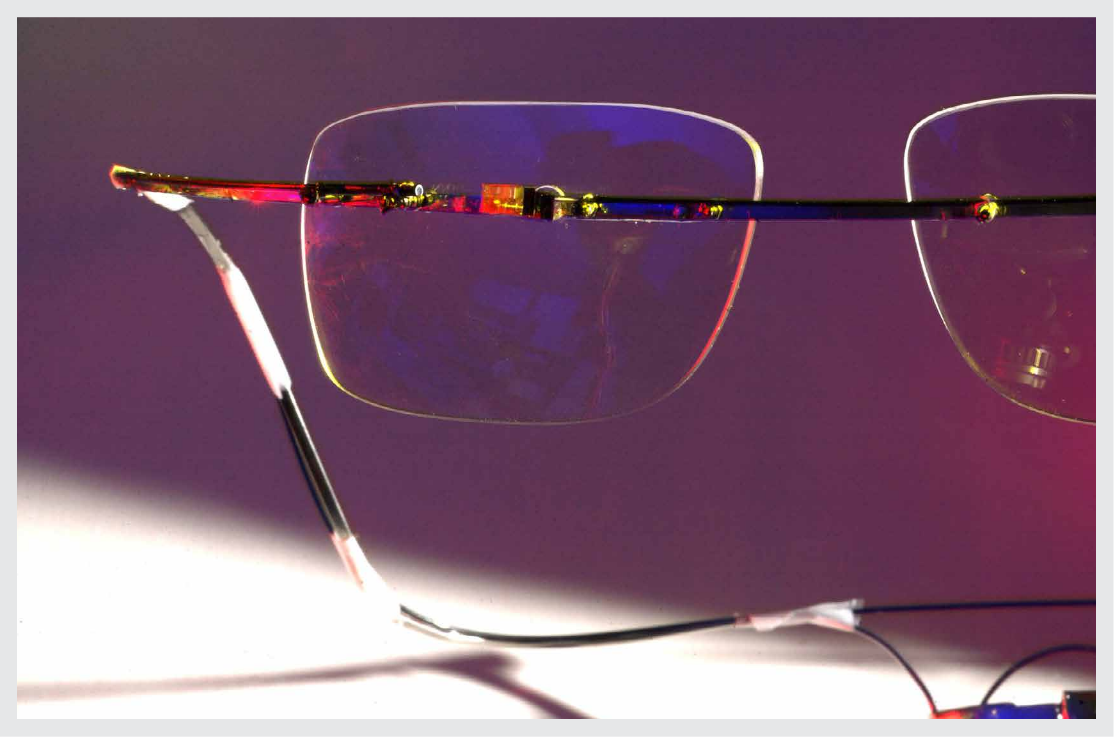


Figure 23.21: Input device for wearable computing: This EyeTap uses a framepiece to conceal the laser light source, which runs along an image fiber optic element, and the camera which runs along another image fiber optic element, the two fibers running in opposite directions, one along the left earpiece, and the other along the right earpiece. This fully functioning prototype of the EyeTap technology has a physical appearance that approximates that of ordinary eyeglasses. The result is a more sleek and slender design.


Figure 23.22: Original wearable computer input devices were inspired by the telegraph key -- This particular telegraph key is a J38 World War II-era U.S. military model.



Morse code was an early form of serial communication, which in modern times is usually automated. In a completely automated teleprinter system, the sender presses keys to send an ASCII data stream to a receiver, and computation alleviates the need for timing to be done by the human operator. In this way, much higher typing speeds are possible.

In simple terms, a keyer is like a keyboard but without the board. Instead of keys fixed to a board like one might find on a desktop, the keys are held in the hand so that a person can press keys in mid air without having to sit at a desk, or the like.

An important application of wearable computing is mediated reality, for which the input+output devices are the sensors and effectors which (a) capture sensory experiences the wearer experiences or would experience; and (b) stimulate these senses. An example is the EyeTap device which causes the eye itself, in effect, to function as if it were both a camera and a display. See Figure 23.23.

An EyeTap having a physical appearance of ordinary eyeglasses, was also designed and built by the author, with materials and assistance provided by Rapp optical; see Figure 23.21.

Figure 23.23: Mann's 'GlassEye™' invention, also known as an EyeTap device, is an input+output device that can connect to a smart phone or other body-borne computer, wirelessly, or by a connection to the AudioVisual ports. A person wearing an EyeTap has the appearance of having a glass eye, or an appearance as if the camera were inside the eye, because of the diverter which diverts eyeward-bound rays of light into the camera system to resynthesize them, typically in laser light. The wearer of the eyetap sees visual reality as re-synthesized from the laser light (computer-controlled laser light source). Pictured here is designer Chris Aimone who collaborated with Mann on this design.

## 23.7 LIFEGLOGGING

From 1994-1996, Steve Mann conducted a Wearable Wireless Webcam experiment where he streamed live video from his wearable computer to and from the World Wide Web, on an essentially 24 hour-a-day basis., For the most part, the wearable computer streamed continuously although the computer itself was not waterproof so it needed to be set aside during showering or bathing. As a personal data capture, Wearable Wireless Webcam raised some new and interesting issues in the capture and archival of a person's entire life from their own perspective. And it also opened up some new ideas such as the roving reporter, where day-to-day living can result in serendipitous capture of newsworthy events. See Figure 23.24 and Joi Ito's chronology of moblogging/lifelogging at the end of this chapter.

In another incident the author was the victim of a hit-and-run. The visual memory of the incident resulted in the arrest and prosecution of the perpetrator.

Wearable Wireless Webcam was at the nexus of art, science, and technology, i.e. it followed a tradition commonly used in contemporary endurance art. For example, it was akin to the living art and endurance art of Linda Montano and Tehching Hsieh who tied themselves to opposite ends of a rope, and remained that way 24 hours a day for a whole year, without touching one another.

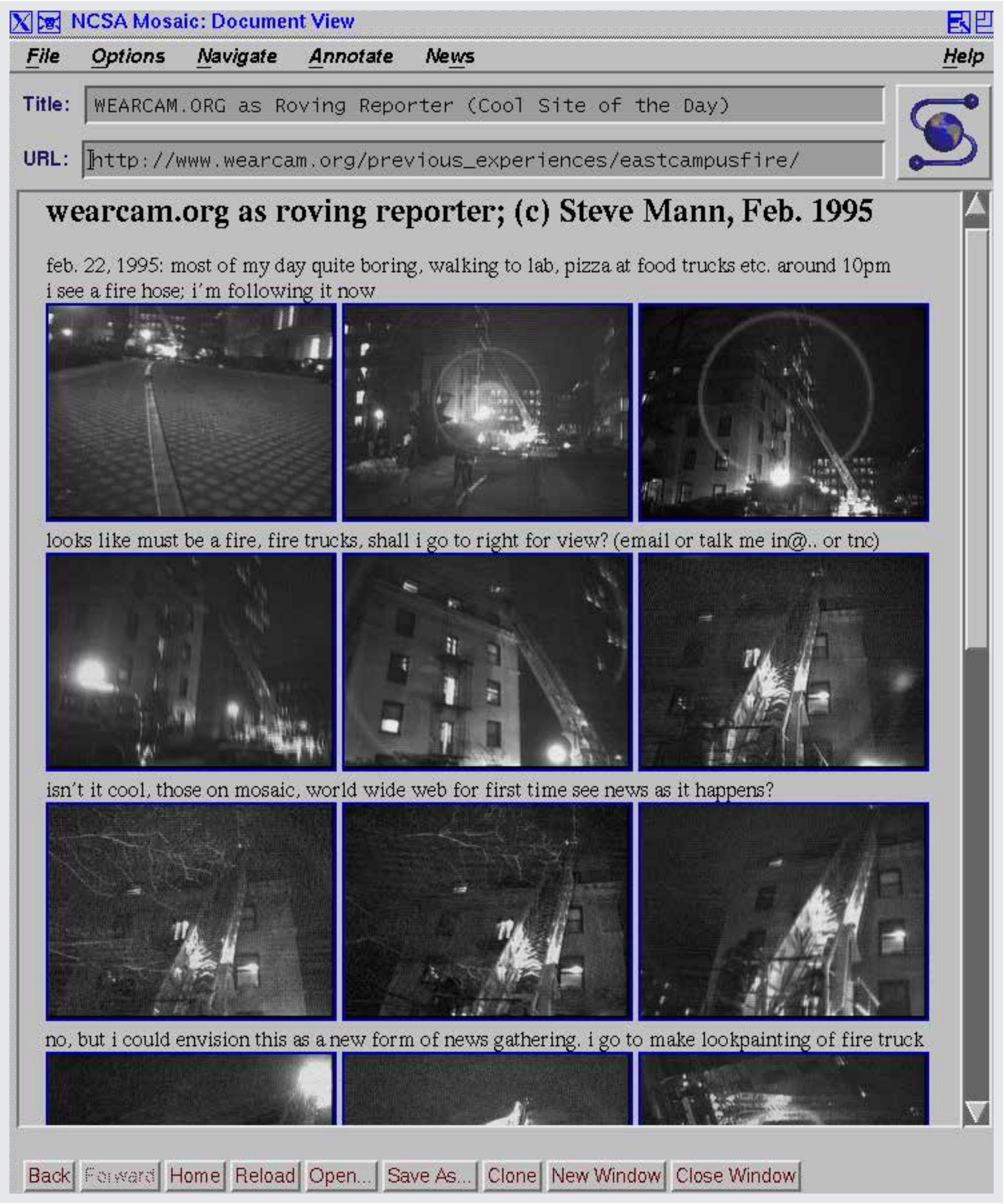



Figure 23.24: Screenshot from Steve Mann's Wearable Wireless Webcam experiment from 1994-1996. Real-time webcast of everyday life resulted in the serendipitous capture of a newsworthy incident. Interestingly the traditional media like newspapers ?had no pictures of the incident, so this is the only photographic record of the incident.

*Courtesy of Steve Mann. Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.*

But it also served a scientific purpose (controlled in-lab experiments are more controlled but make a trade-off between external validity versus internal validity), and an engineering or technical purpose (inventing new technologies, etc.).

One interesting by-product of Wearable Wireless Webcam was the concept of lifeglogging, also known as cyborGLOGGING, glogging, lifelogging, lifecasting, or sousveillance.

The word “surveillance” derives from the French words “sur”, meaning from above, and “veiller” meaning “to watch”. Surveillance therefore means “watching from above” or “overwatching” or “oversight”. While much has been written about surveillance and the relative balance between privacy and security (i.e. some people arguing for more surveillance and others arguing for countersurveillance), this argument is one-dimensional in that it functions like a one degree-of-freedom “slider” to choose more or less surveillance. But sousveillance (“sous” is French for “from below” so the English word would be “undersight”) has recently emerged as an alternative.

Sousveillance refers to the recording of an activity by a participant in the activity, typically by way of small wearable or portable personal technologies (Mann et al 2003, Mann 2004, Dennis 2008, Baikr 2010, Deirdre 2009, Thompson 2011, Brin 2011).

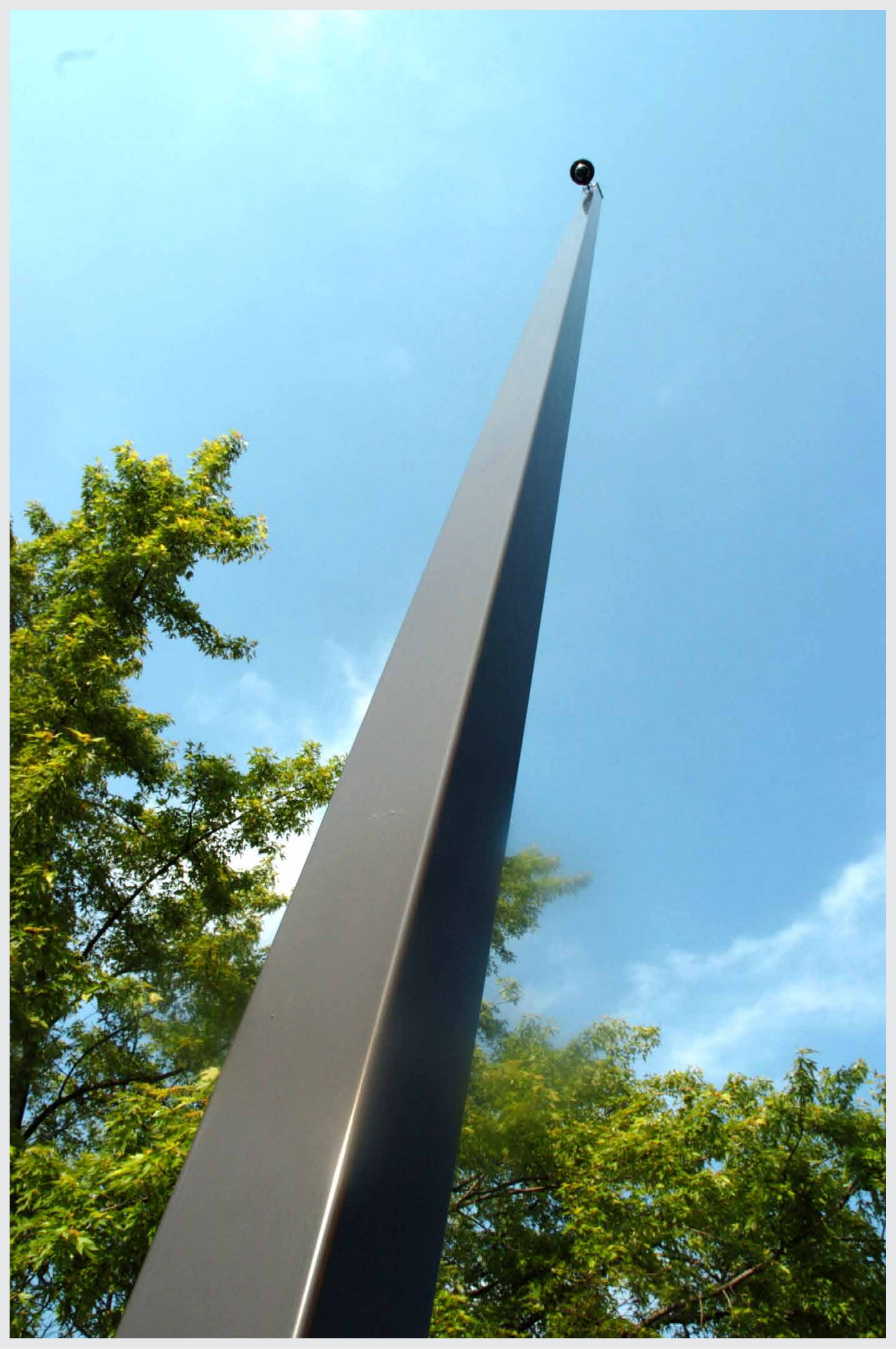




Figure 23.25 A-B: Leftmost, a surveillance dome camera atop a lamp post serves as an "eye-in-the-sky" watching down on a parking lot. Rightmost: a surveillance dome as a necklace has a fisheye lens and various physiological sensors. Sensor camera designed and built and photographed by Steve Mann 1998. Mann presented this invention to Microsoft Corporation as the Opening Keynote at ACM Multimedia's CARPE in 2004 (http://wearcam.org/carpe/).

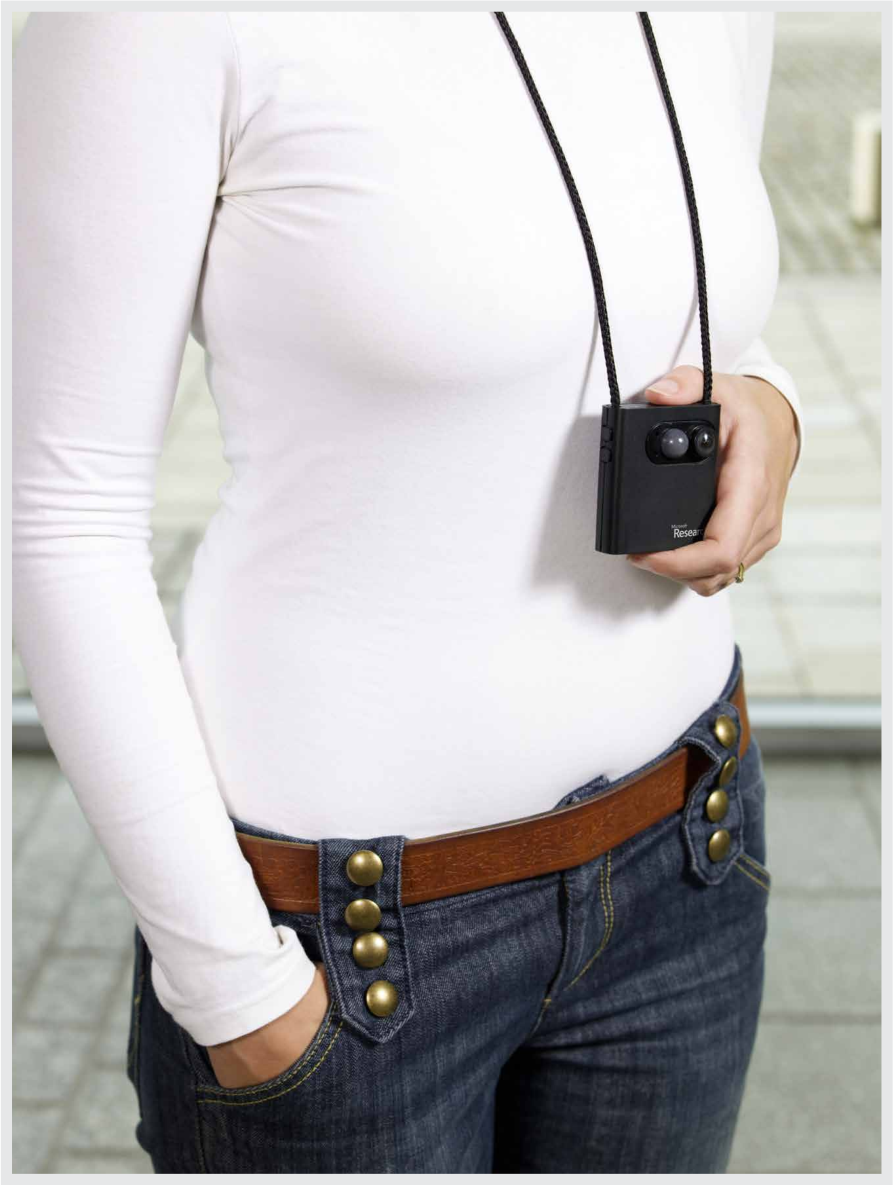





FIGURE 23.26 A-B: Around 2005, Microsoft built and researched their SenseCam prototype - a version of the neckworn camera. It was commercialized in 2009 (licenced to Vicon) and is now available as a product called Vicon Revue.

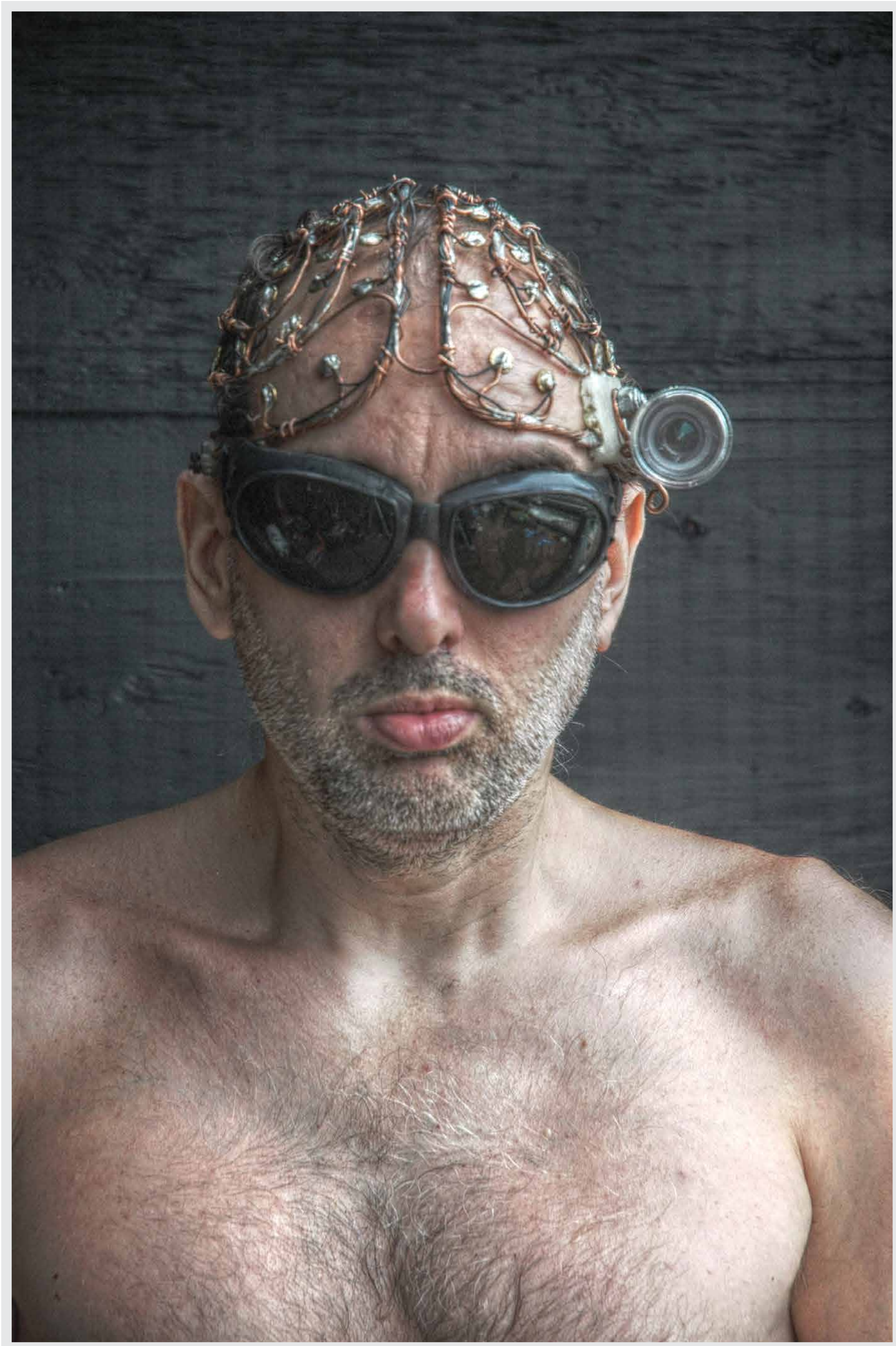
While surveillance and sousveillance both generally refer to visual monitoring (i.e. “veiller” being “to watch”), the terms also denote other forms of monitoring such as audio surveillance or sousveillance. In the audio sense (e.g. recording of phone conversations) sousveillance is referred to as “one party consent”.

## 23.8 FUTURE DIRECTIONS AND UNDERLYING THEMES

### 23.8.1 Cyborgs, Humanistic Intelligence and the reciprocal relationship between man and machine

The wearable computer can provide many benefits, such as assistive technologies to help people see better, remember better, and function better, e.g. for the elderly to age gracefully, or for those with Alzheimer’s disease to be able to remember and recognize names and faces.

One project, the author’s Mindmesh, enables the blind to see, and people with visual memory impairment to remember and recall visual subject matter. The Mindmesh comprises a permanently attached skull cap with a combination of implantable and surface electrodes, as well as a mesh-based computing architecture in which individual processors are each responsible for eight electrodes. The Mindmesh, still in its early development stages, is evolving toward an apparatus that allows the user to plug various sensory devices “into their brain” in a sense. So a blind person will be able to plug a camera into their brain, or an Alzheimer’s patient will be able to attain a form of autoassociative memory. See Figure 23.27 A-B



*Courtesy of Steve Mann. Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.*

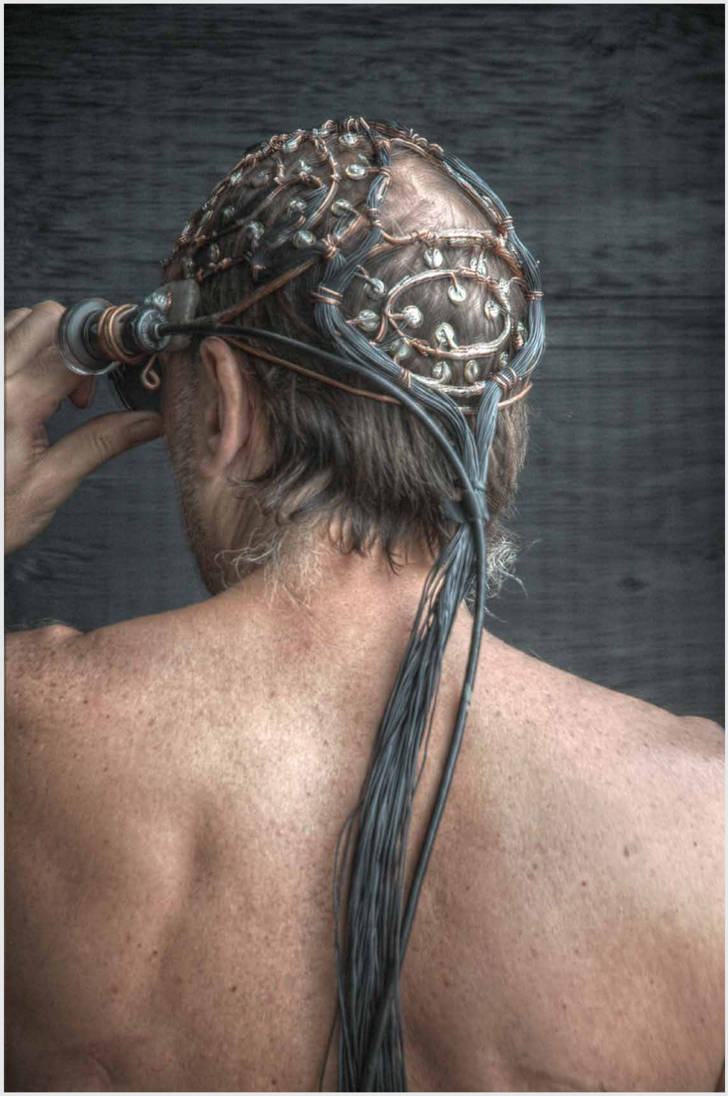




Figure 23.27 A-B: The Mindmesh is a mesh-based computing architecture currently under development, to allow various sensors and related devices to be "plugged into the brain". Some variations of the Mindmesh can be permanently attached, and are ruggedized to withstand the rigours of life, e.g. running through fountains or jumping into the ocean, etc. The author wishes to thank Olivier Mayrand, InteraXon, and the OCE (Ontario Centres of Excellence) for assistance with this work.

The Visual Memory Prosthetic (VMP) is thus combined with the new computational seeing aid, which can thusly capture a cyborglog of a person's entire life and hopefully in the future be able to index into it. This is part of the author's "Silicon Brain" project in which the Mindmesh indexes into an autoassociative memory to assist persons with sensory integration disorder, or the like. As we replace more of our mind with external memory, these memories become part of us, and our own personhood.

Businesses and other organizations have a legal obligation not to discriminate against persons with special needs, or the like, or to treat persons differently depending on such technologies. As we see the widespread adoption of technologies like Mindmesh, which, essential to their functioning as memory aids, must capture, process, and retain data, may be interpreted as making recordings. The "Silicon Brain" of the Mindmesh thus asks the question "is remembering recording?". As more people embrace prosthetic minds, this distinction will disappear. Businesses and other organizations have a legal obligation not to discriminate, and will therefore not be able to prevent individuals from seeing and remembering, whether by natural biological or computational means.

But we don't have to even wait for the future widespread adoption of the Mindmesh to observe culture in contention. As mentioned earlier, smartphones are the precursor to full-on wearable computing, and their proliferation has already brought forth this very issue.

### 23.8.2 Privacy, surveillance, and sousveillance

Surveillance is an established practice, and while controversial, much of the controversies have been worked out and understood. Sousveillance, however, being a newer practice, remains, in many ways, yet to be worked out.

The proliferation of camera phones itself has even resulted in numerous cases in which police and security guards have been caught in wrongdoing. Also, there have been numerous cases where police and security guards have tried to destroy evidence captured by ordinary citizens. In one case, a man named Simon Glik was arrested for recording actions of police officers on video. However, The United States Court of Appeals ruled in favor of Glik, and after finding no wrong doing on his part, the courts found that the officers violated Glik's first and fourth amendment rights[1].

These controversies will not go away, though, as the line between remembering and recording, and between the eye and camera, blur. Many department stores and other establishments have signs up prohibiting photography and prohibiting cameras (see Figure 23.28 A-B-C), but they also have 2-dimensional barcodes designed to be read by patrons using their smartphones (see Figure 23.29 A-B). Thus they simultaneously encourage patrons to take photographs and prohibit patrons from taking photographs.


*Courtesy of Steve Mann. Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.*

1. See the full details at http://www.citmedialaw.org/sites/citmedialaw.org/files/10-1764P-01A.pdf


*Courtesy of Steve Mann. Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.*




Figure 23.28 A-B-C: Signs that say "No video or photo taking" and "NO CELL PHONE IN STORE PLEASE!" are commonplace, yet people are relying more and more on cameras and cellphones as seeing aids (hand-held magnifiers to help them read the very signs that prohibit their use, for example), and to access additional information that will help them make a purchase decision.


*Courtesy of Steve Mann. Copyright: .*


*Courtesy of Steve Mann. Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.*

Figure 23.29 A-B: "SCAN ME... Use your smartphone to scan this QR code..." says the box in a store where cellphones and cameras are forbidden.


Figure 23.30: The irony of treating cameras and cellphones as contraband in semi-public places is that this trend seems to come around the same time as the proliferation of CCTV surveillance cameras. In the future, when a security guard demands a patron remove their electric eyeglasses, the guard may be liable when the patron trips and falls. The authority of the guard does not extend to mandating eyeglass prescriptions of their customers.



### 23.8.3 Copyright and ownership: Who do memories, information and data belong to?

This potential controversy extends to content. For example, the trend toward licensed software leads to a situation where licenses expire and the computer stops working when the license fee is not paid. When a computer program is helping someone see, a site license becomes a sight license. As our bodies and computing increasingly intersect and intertwine, as in the case of Wearable Computing, we must ask ourselves if we really want to live in a society where “your pacemaker firmware license is about to expire, please insert your credit card to continue living”. (Mann 2003). This theme was the topic of an art installation at San Francisco Art Institute (SFAI), in response to their request for an exhibit on wearable computing. See Figure 23.31


Figure 23.31: Wearable computing exhibit at San Francisco Art Institute 2001 Feb. 7th. This exhibit comprised a chair with spikes that retract for a certain time period when a credit card is inserted to purchase a seating license.


Figure 23.32: Poster for the wearable computing exhibit at San Francisco Art Institute 2001.

*Courtesy of Steve Mann. Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.*

By its very nature, wearable computing evokes a visceral response, and will likely fundamentally change the way in which people live and interact. In the future, devices that capture our lifelong memories, and share them in real-time, will be commonplace and worn continuously, and perhaps even permanently implanted. As an example, the author has invented and filed a patent for an artificial eye that provides people with vision in one eye, an implantable eye to see stereo using a crosseyetap. Various filmmakers have approached the author requesting help embodying this invention. See for example Figure 23.33.


Figure 23.33: Occular implant artificial eye camera invented by Steve Mann (Canadian Pat.2313693), and built by Rob Spence and others, in collaboration with Mann. The artificial eye has a camera built into it for persons with vision in only one eye; the eye may thus function as a wearable wireless webcam and cyborglogging device, and hopefully soon in the future as a vision replacement. With the computer and camera implanted fully inside the body, some people are able to stream live video without having to wear anything. The apparatus has the appearance of a normal eye, yet provides sousveillance and cyborglogging (today) and may provide vision replacement tomorrow.

## 23.9 WHERE TO LEARN MORE

*Mann, Steve (1997): Wearable Computing: A First Step Toward Personal Imaging. In IEEE Computer, 30 (2) pp. 25-32*

*Mann, Steve (2001): Intelligent Image Processing. Wiley-IEEE Press*

*Mann, Steve (1996): Smart Clothing: The Shift to Wearable Computing. In Communications of the ACM, 39 (8) pp. 23-24*

*Mann, Steve and Niedzviecki, Hal (2001): Cyborg: Digital Destiny and Human Possibility in the Age of the Wearable Computer. Doubleday of Canada*

- Marketing Wearable Computers to Consumers
- IEEE Special Issue on Wearable Computing and Humanistic Intelligence
- Editor's Blog Sousveillance: Wearable Computing and Citizen "Undersight"
- Bootstrap Knowledge Web
- Barfield, Woodrow and Caudell, Thomas (eds.) (2001): *Fundamentals of Wearable Computers and Augmented Reality*. CRC Press

### 23.9.1 Wearable Computing Conferences

The first wearable computing conferences were:

- The International Conference on Wearable Computing (ICWC) and
- The International Symposium on Wearable Computing (ISWC)

### 23.9.2 Historical chronology of moblogging, also known as cyborglogging, lifeglogging, lifelogging, lifecasting, and the like

The following is a "chronology of articles, events and resources, About moblogging", written by Joi Ito:

- February 1995 - wearcam.org as roving reporter Steve Mann
- January 4, 2001 4:16p - Stuart Woodward first posts from his cellphone on Stuart Woodward's LiveJournal using J-Phone, Python and Qmail
- January 6, 2001 15:09 - First reported posting from under the sea
- Thursday, March 1, 2001 - First post by SMS to David Davies' SMSblog and his announcement on the Radio Userland Support List
- January 11, 2002 - Radio Userland released with mail to post feature
- January 13, 2002 - A text post from an Docomo P503i to Al's Radio Weblog
- February 18, 2002 - Hiptop Photo Gallery started by Michael Morrissey (using procmail and PERL script later improved by Dave Bort who shared it to people including Mike Popovic who started Hiptop Nation)
- February 2002 - Justin Hall posts pictures and text to Jiqoo.com using a a J-Phone and Brian Hooper's code. (Link is now dead)
- Fisher, Scott S., " An Authoring Tool Kit for Mixed Reality Experiences", International Workshop on Entertainment Computing (IWEC2002): Special Session on Mixed Reality Entertainment, May 14 -17 2002, Tokyo, Japan
- Summer 2002 - Kuku Nipernaadide - Estonian moblog started by Peeter Marvet
- 2001 - Howard Rheingold coins the word "Smartmobs"
- November 5, 2002 - Adam Greenfield coins the term "Moblogging"
- October 1, 2002 - T-Mobile Sidekick launch moblog (Danger internal) 149 pictures in 24 hours

- Friday October 4, 2002 9:15a - Hiptop Nation created by mikepop
- October 31, 2002 - Hiptop Nation Halloween Scavenger Hunt
- November 21 2002 - The Feature article From Weblog to Moblog by Justin Hall
- Tuesday, November 26, 2002 - Stuart Woodward's first image posted from a cell phone
- November 27, 2002 10:41p - Joi Ito's Moblog (using attached image mail to MT)
- December 4, 2002 - Milano::Monolog blogs text from mobile phones (Japanese)
- December 12, 2002 - Guardian article Weblogs get upward mobility
- December 31, 2002 - New Year's Eve Moblog Blog-Misoka by JBA
- January 8, 2003 - electricnews.net Start-up marries blogs and camera phones
- January 8, 2003 - The Register Start-up marries blogs and camera phones
- January 9, 2003 - Robert announces PhoneBlogger

## 23.10 COMMENTARY BY KATINA MICHAEL AND M. G. MICHAEL

How to cite this commentary in your report

### Katina Michael

*© Katina Michael*

Katina Michael is the IEEE Technology and Society Magazine editor-in-chief. She is the author of Innovative Automatic Identification and Location-Based Services: from Bar Codes to Chip Implants (2009) and has hosted six workshops on the Social Implications of National Security. Michael (MIEEE'04, SMIEEE'06) holds a Doctor of Philosophy in Information and Communication Technology (ICT)...

Katina Michael
Katina Michael is a member of The Interaction Design Foundation

## M. G. Michael

© *M. G. Michael*

M.G. Michael coined the term "überveillance" in 2006 to denote omnipresent electronic surveillance embedded beneath the skin. The term was entered into the official dictionary of Australia, the Macquarie Dictionary in 2008. Michael received a PhD from the School of Theology at the Australian Catholic University in 2003 in Brisbane, Queensland, and a Master of Arts Honors from the Sch...

M. G. Michael
M. G. Michael is a member of The Interaction Design Foundation

### 23.10.1 About Steve Mann

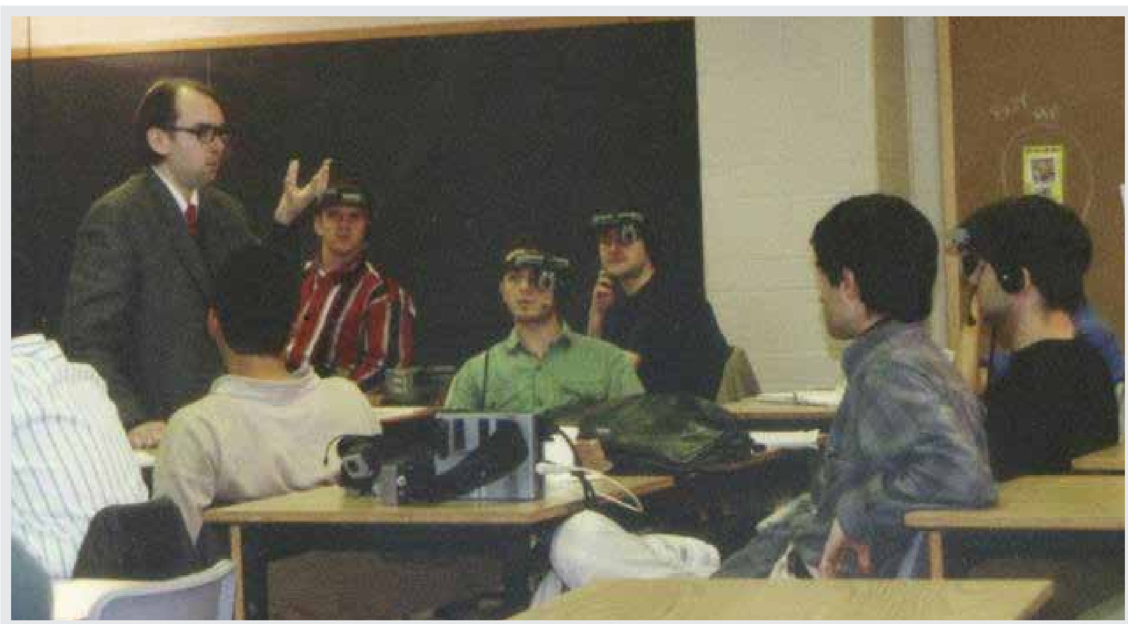
In Professor Steve Mann - inventor, physicist, engineer, mathematician, scientist, designer, developer, project director, filmmaker, artist, instrumentalist, author, photographer, actor, activist - we see so much of the paradigmatic classical Greek philosopher. I recall asking Steve if technology shaped society or society shaped technology. He replied along the lines that the question was superfluous. Steve instead pointed to *praxis*, from which all theory, lessons or skills stem, are practiced, embodied and realized. Steve has always been preoccupied by the application of his ideas into form. In this way too, he can be considered a modern day Leonardo Da Vinci.

It is not surprising that Professor Mann was awarded the *2004 Leonardo Award for Excellence* (Leonardo, 2004). In his winning article he presented “Existential Technology” as a new category of in(ter)ventions and as a new theoretical framework for understanding privacy and identity (Mann, 2003). At the time, Mann had written more than 200 research publications already, and was the keynote speaker at numerous industry symposia and conferences. His work had also been shown in museums globally, including the Museum of Modern Art in New York, the Stedelijk Museum in Amsterdam and the Triennale di Milano (Quillian, 2004).

I embarked on my PhD in 1997, the same year in which Steve graduated with a PhD in Media Arts and Sciences from MIT under the supervision of Professor Rosalind Picard and MIT Media Lab creator and director Professor Nicholas Negroponte. I remembered being amazed by the research Steve was engaged in, particularly his insights into wearable computing, and thinking all at once what incredible uses the technologies he was developing could have but also what they might mean in terms of the social implications. At that time I was working for Nortel Networks as a network planner and strategically positioning big pipes throughout the world in anticipation of the big data that was coming through IP-based applications. Few, however, could possibly have imagined that people would be willingly creating lifelogs (or cyborg logs) through the act of glogging (PCMag, 2012), another Mann discovery, and uploading them in real-time through wireless technology, every minute of every hour of every day (Mann, 1995). 4G will make glogging even easier. Presently, there are over 165,000 gloggers at http://glogger.mobi.

### 23.10.2 Corresponding with Professor Mann about Sousveillance in the Educational Context

At the beginning of 2009, my close collaborator Dr MG Michael and I decided to explore the idea of glogging, inspired by correspondence with Steve on his notion of existential education (ExistEd), Figure 23.1, first officially demonstrated in 1998 (Mann, 1998). We asked our class of 163 undergraduate and postgraduate students in a compulsory computer ethics course to take part in some personal field work through the use of glogger.mobi. Examples from this class can be found in Table 1. We wanted to see sousveillance acted out before us, what its limits were, if any, and we wanted to attempt it without having sought prior Human Research Ethics Committee approval.



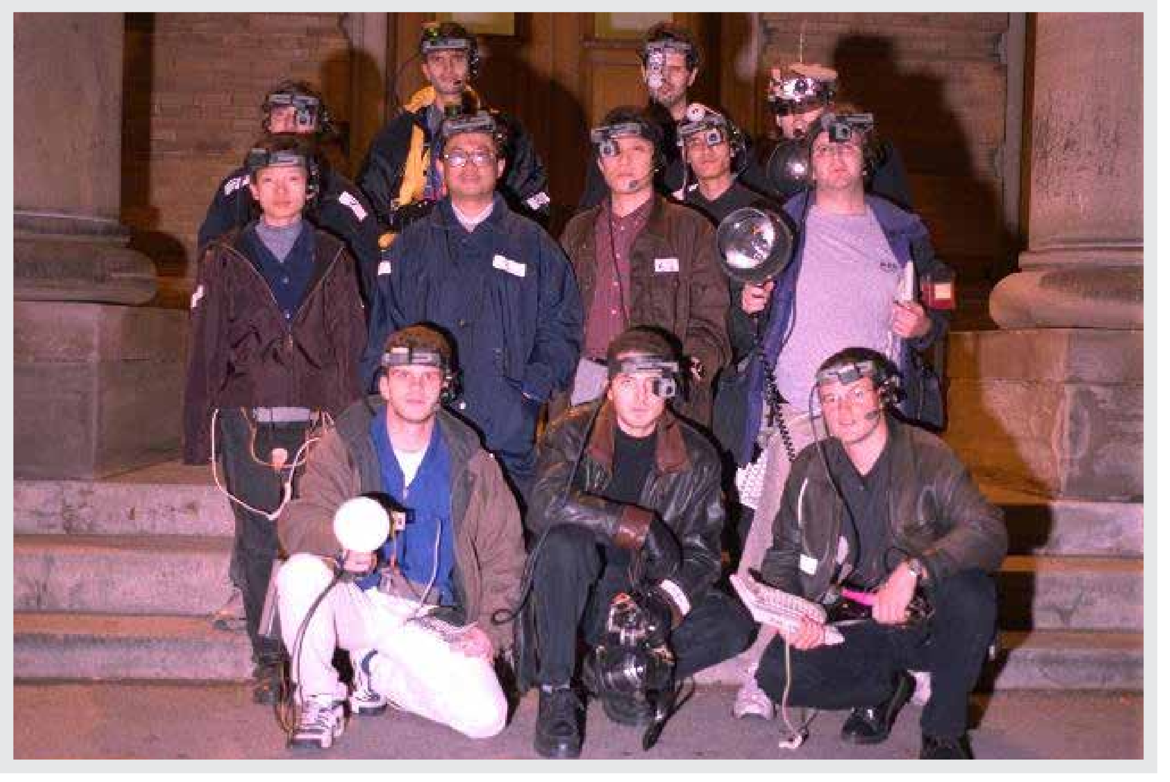




FIGURE 23.1 A-B-C: Professor Steve Mann's Existential Learning = "Learn by being [a photoborg]"

| Storyboard Title | Creator | Date Uploaded |
| --- | --- | --- |
| Computer Ethics Sock Puppet Theatre | kshuntley | 20 April 2009 |
| Nanny Cams | watguy | 20 April 2009 |
| In the 'hood: Identity theft from your home | randomisa-tions | 20 April 2009 |
| Photography and privacy | Jiang | 20 April 2009 |
| Cameras and Privacy | ls013 | 23 April 2009 |

| | | |
|---|---|---|
| Invasion of Location Privacy | akarin | 23 April 2009 |
| Nanotech Future | Jeesper | 23 April 2009 |
| Mobile phone privacy | ml733 | 19 April 2009 |
| Identity Fraud | Hooka69 | 20 April 2009 |
| RFID Issues | tlc91 | 20 April 2009 |
| DNA | seven3one | 20 April 2009 |
| Health Insurance | cjal073 | 20 April 2009 |
| Bluetooth - The phone is mightier than the handgun | rt902 | 4 May 2009 |
| Australian Government 'Cleanfeed' Internet Filtering Scheme | ha766 | 24 April 2009 |
| Life inside a camera | skt999 | 20 April 2009 |
| 2 guys one camera | ads32 | 19 April 2009 |

TABLE 23.1: Storyboards Created by University of Wollongong Students during the Course IT and Citizen Rights, Session 1, 2009.

It was interesting to observe that in our class of 163 there were:

- a handful of students who claimed their parents would not allow them to upload their own image, or images of others online;
- about fifteen students who refused to take any photographs themselves using any camera device and instead downloaded images from the public domain (many of them copyrighted);
- a handful of students who did not wish to participate in a public glog without any privacy controls but who would have otherwise considered participation;

- dozens of students who thought it was inappropriate to film others without asking their permission in a public setting even if they were in the point of view or fellow participants in events they were engaged in;
- about a dozen students who thought it would be better to use cartoon characters or puppets instead of humans in their storyboards;
- a handful of students who did not wish to disclose their identity and so wore a hood or hat or high collar;
- about a dozen students (mostly internationals) who removed their storyboards the same day assessment results were returned to them; and
- a handful of students who possibly went too far and filmed or photographed sensitive data such as Automatic Teller Machines (ATMs) at point blank, or other very personal activities they were engaged in.

We cannot claim that in all cases our students were 'learning by being' (a step beyond 'learning by doing'), but some did become true photoborgs, whilst others took on the persona of a photoborg, even if it was for a few short weeks. It takes nerve for someone to actually wear a camera, not just carry it, to admit to it recording when questioned, and to cope with the responses that that kind of activity might provoke in a setting like a regional centre in Australia. But as Steve plainly emphasizes, "[w]hat really matters, much more than whether the technology is implanted, worn, carried, or non-existent, is the degree to which the educational paradigm embodies an epistemology of personal choice, and the metaphysics of personal freedom, growth, and development" (Mann, 2006). Furthermore, Mann writes about deconstructionist learning: "As a "cyborg" in the sense of long-term adaptation, body-borne technologies, etc., one encounters a new kind of existential self-determination and mastery over one's own destiny, that can be learned, in the postmodern (posthumanism) context one might think of as the "cyborg age" in which many of us now live."

Increasingly, photoborgs are now everywhere, and as they increase in numbers over the next decade, comfort levels of photoborg presence will also likely increase as it becomes commonplace. However, there are still laws for instance, that are in direct conflict with photoborgology. See, for example, the Surveillance Devices Act in the state of Western Australia, in Australia (WA, 1998):

SURVEILLANCE DEVICES ACT 1998 - SECT 6

- 6. Regulation of use, installation and maintenance of optical surveillance devices
    - (1) Subject to subsections (2) and (3), a person shall not install, use, or maintain, or cause to be installed, used, or maintained, an optical surveillance device
    - (a) to record visually or observe a private activity to which that person is not a party; or
    - (b) to record visually a private activity to which that person is a party.
    - Penalty:
    - (a) for an individual: $5 000 or imprisonment for 12 months, or both;
    - (b) for a body corporate: $50 000.

As Roger Clarke (2012b) has pointed out, this act: "seems to mean that, although you can audio-record your own conversations with other people, you can't **video**-record them... That has serious implications for sousveillance, i.e. the use of surveillance by the less powerful, when dealing with the more powerful".

It was in preparation for our class that we discussed the ultimate trajectory of wearables and implantables with Steve, finding his work on the existentiality axis (Figure 23.2) to be critical (Mann, 2001; Mann, 1998). It just so happened at the time we were corresponding, that Steve was working on a project with the Eyeborg Man, Rob Spence (Spence, 2008) and we were in full preparation to host the International Symposium on Technology and Society (ISTAS10) at the University of Wollongong which had as one of its major themes microchip implants for humans (UOW, 2010; Michael, 2011). Researcher Mark Gasson and Mr Amal

Graafstra, both bearers of chip implants spoke on their experience of being radio-frequency identification (RFID) implantees at the symposium.

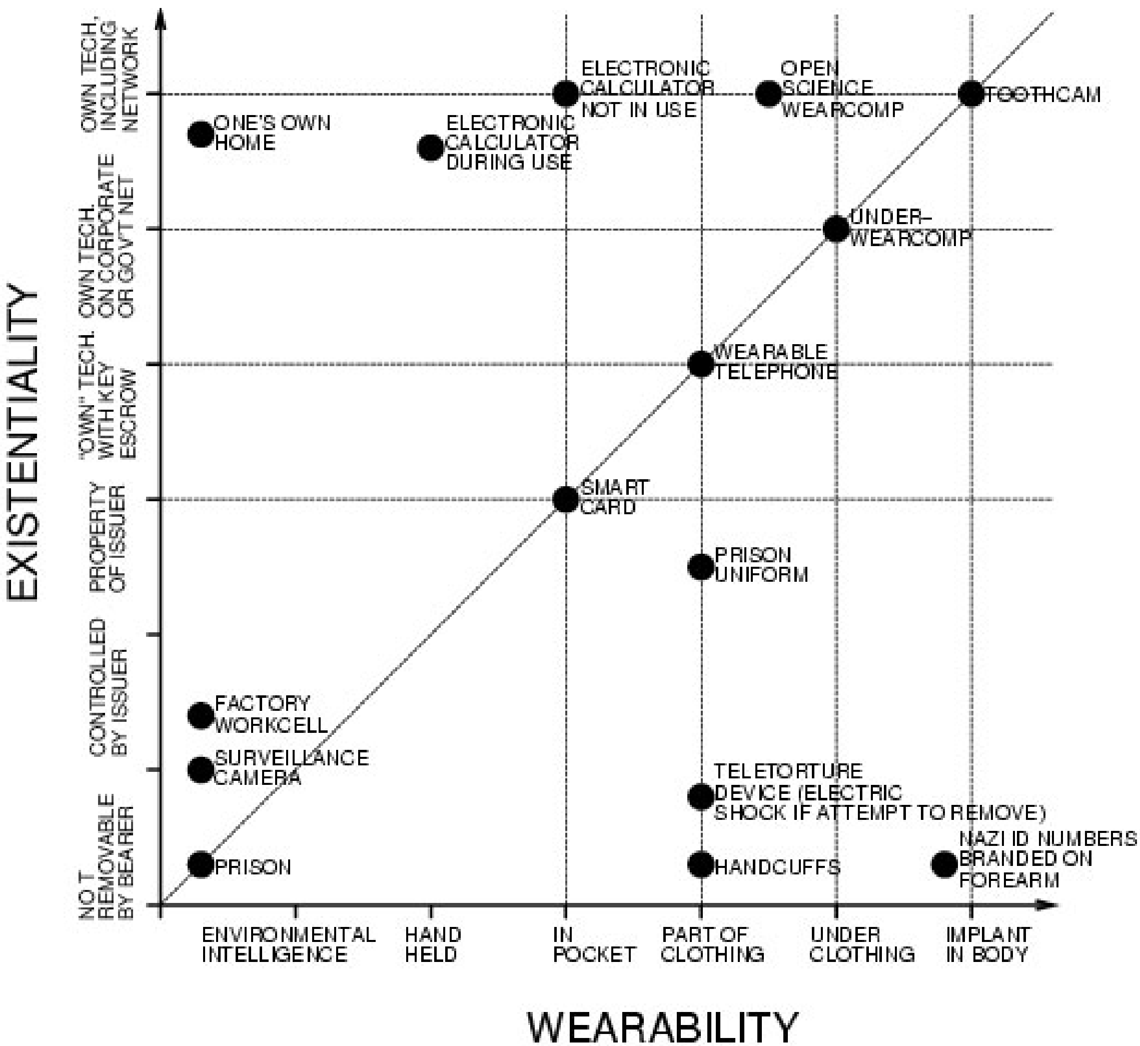



FIGURE 23.2: Wearability/Portability versus Existentiality.



It was during this time that the interface between sousveillance and überveillance began to emerge. On the one hand, you had camera technologies that people wore to conduct surveillance "from below", and on the other we had proposed in 2006 that embedded systems, such as implantables, would one day do the surveilling "from within". It was in the Eyeborg's 'implantable' camera that sousveillance came face-to-face with überveillance (Michael and Michael, 2009). In Figure 23.3, the various veillances are depicted in a triquetra by Mr Alexander Hayes.

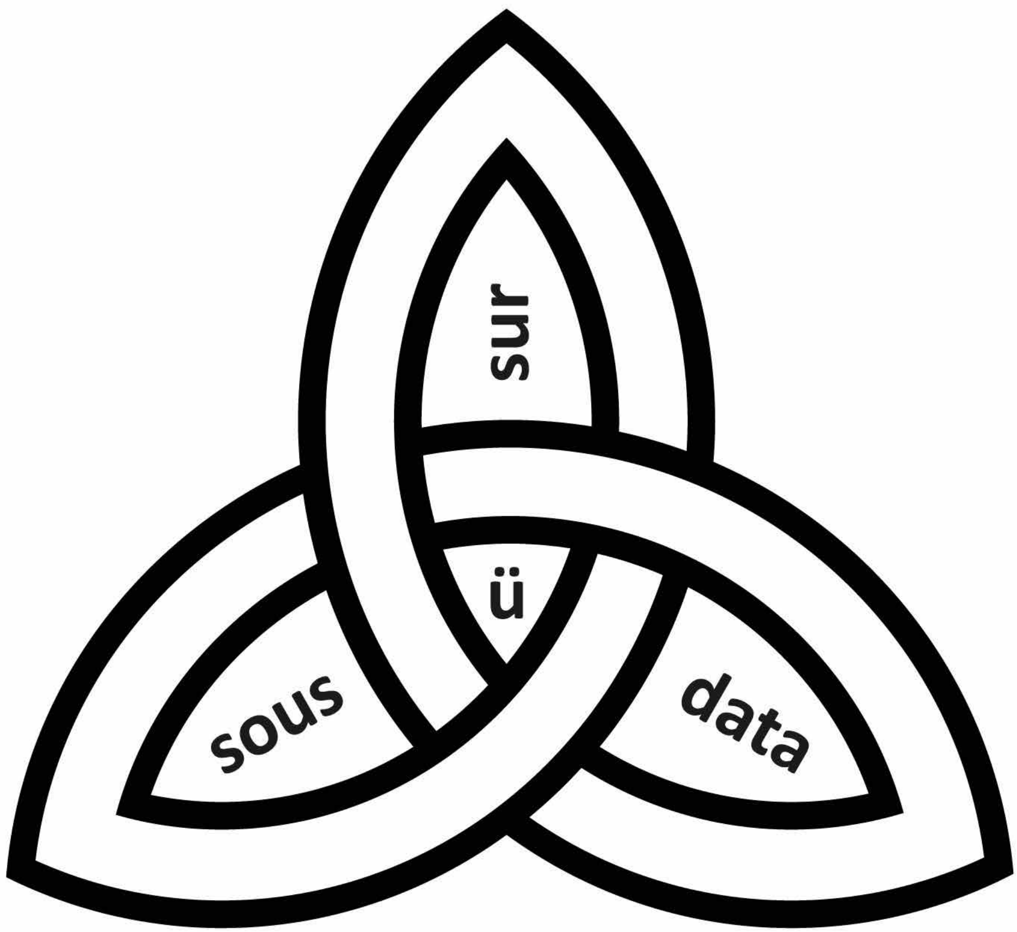



FIGURE 23.3: The ?berveillance Triquetra by Mr Alexander Hayes.

### 23.10.3 Sousveillance Outside the Context of Existential Education

Taken away from the context of learning and reflection, a wearable (or implantable) camera worn by any citizen carries with it significant and deep personal and societal implications. The photoborg may feel entirely free, a master of his/her own destiny; they may even feel safe that their point of view is being noted for re-use, if needed at a later time. Indeed, the power the photoborg presumes when they put on the camera or bear the implant, can be considered even more powerful than the traditional CCTV overhead gazing in an unrestricted manner. But no matter how one looks at it, others will inevitably be in the field of view of the wearer or bearer of technology, and unless these fellow citizens also become photoborgs themselves, there will always be inequality. Professor Mann's sousveillance carries with it huge political, educational, environmental and spiritual overtones. The narrative which informs sousveillance is more relevant today due to the proliferation of new media, than ever before. But wherein sousveillance grants the citizen the ability to combat the powerful using their own strategic game, it also grants other citizens the ability to put on the guise of the powerful. Sousveillance is here in the eye of the beholder, the one wearing the camera. In the end it comes down to lifeworld and context and stakeholder types. What we all agree on however, is the pervasiveness of the camera, that sees everything, hears everything but comes endowed with obvious limitations such as the potential for the impairment of data, through data loss, data manipulation, or misrepresentation.

MG and Steve had similar conceptions of where the surveillance capability of the powerful is going, expressed so eloquently during the Singularity Summit in San Francisco where Steve described his Ladder Theory (Mann, 2010). MG Michael likewise referred to the idea of the "axis of access" in 2010, which Steve noted would be more correct if written "axes of access". It was unsurprising to us, in conducting research for this article on Steve's wearable computing history, that we stumbled across the following *Wired* article written in 2003 by Shachtman:

"The Pentagon is about to embark on a stunningly ambitious research project designed to gather every conceivable bit of information about a person's life, index all the information and make it searchable... The embryonic LifeLog program would dump everything an individual does into a giant database: every e-mail sent or received, every picture taken, every Web page surfed, every phone call made, every TV show watched, every magazine read... All of this -- and more -- would combine with information gleaned from a variety of sources: a GPS transmitter to keep tabs on where that person went, audio-visual sensors to capture what he or she sees or says, and biomedical monitors to keep track of the individual's health... This gigantic amalgamation of personal information could then be used to "trace the 'threads' of an individual's life."

It simply goes to show how any discovery can be tailored toward any ends. Glogging was meant to sustain the power of the individual, to enable growth, maturity and development in the person. Here, it has been hi-jacked by the very same stakeholder it was created to gain protection from. Many would ask are we playing into the hands of such initiatives as DARPA's Lifelog program by researching sousveillance and überveillance. The answer to this is not difficult- the natural trajectory of these emerging technologies would have propelled us there regardless. Arthur (2009, p. 15) speaks of an evolution of technology which is "the process by which all objects of some class are related by ties of common descent from the collection of earlier objects." If it were not Steve Mann, then someone else would have at some given point in time discovered the true capabilities of sousveillance -- better for it to have been Mann who embraces genuine discussion on issues related to privacy, identity, and human rights.

### 23.10.4 International Workshop in Recognition of Steve Mann's Sousveillance Research

Mid-way through 2011, MG and I decided we would host an international workshop on Sousveillance and Point of View (POV) Technologies in Law Enforcement as our sixth workshop in the *Social Implications of National Security* series (Michael and Michael, 2012b) which started as an initiative under the Australian Research Council's (ARC) Research Network for a Secure Australia (RNSA). The workshop was hosted on the 22 February 2012, exactly 17 years after Mann uploaded his wearable webcam images of MIT's east campus fire as a "roving reporter." But rather than just focusing on sousveillance, the workshop also emphasised the rise of crowd-sourced sousveillance, citizen rights, body-worn technologies and just-in-time policing. The workshop investigated the use of sousveillance by law enforcement for evidence-based gathering as well as its use against law enforcement by everyday citizens. The ways in which we have witnessed the proliferation of overt and covert surveillance technologies has set the stage for the re-evaluation of existing laws and practices.

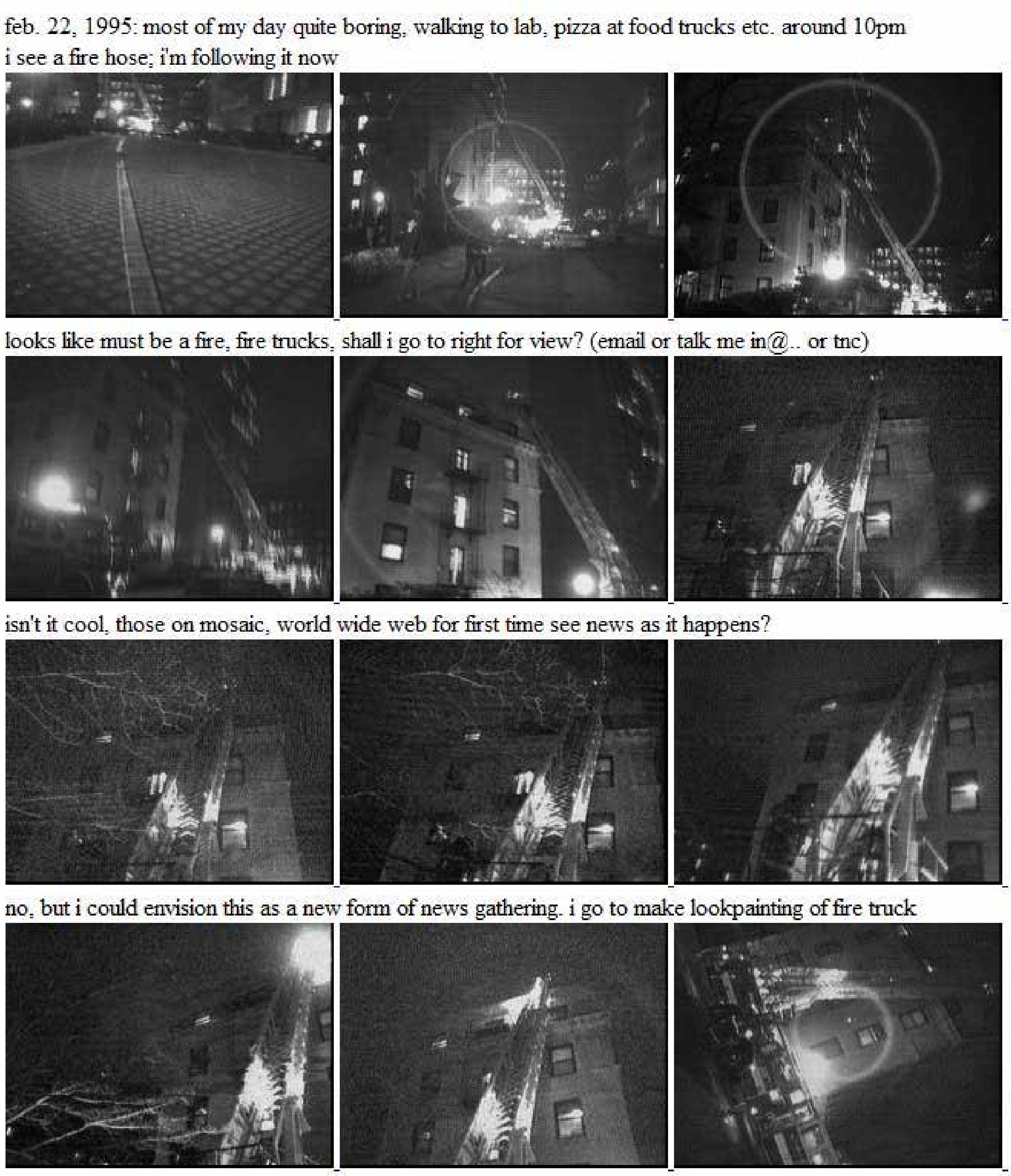



FIGURE 23.4: The International Workshop on Sousveillance and Point of View (POV) Technologies in Law Enforcement was held exactly 17 years after Mann uploaded his wearable webcam images of MIT's east campus fire as a "roving reporter."

Over 50 delegates attended (Figure 23.5) including former Privacy Commissioners of Australia, prosecutors and barristers of the high court, members of the Queensland and Victorian police, private investigators, spy equipment vendors (Figure 23.6), National Information and Communications Technologies Australia (NICTA) representatives, educational technologists (Figure 23.7), the Commissioner for Law Enforcement Data Security, members of the Australian Privacy Foundation, artists (Figure 23.8) and academics from across the country. A highlight of the workshop was the attendance of well-known Canadian sociologists Professors Kevin Haggerty and David Lyon, who gave a keynote address and invited paper (Figure 23.6). Professor Haggerty spoke on the 'Monitoring of Police by Police' and Professor Lyon on the concept of the 'Omniscient Gaze' (Bradwell, 2012) (Figure 23.9).

Figure 23.5: Delegates at the International Workshop on Sousveillance in Australia.

FIGURE 23.6: DUSS Pty Ltd. Demonstrating the eWitness Wearable Camera.

Figure 23.7: Mr Alexander Hayes speaking about Professor Steve Mann and sousveillance during his presentation on POV and Education.

Figure 23.8: Mr Tim Burns. Western Australian Artist speaking at the Workshop.


Figure 23.9: Professor Kevin Haggerty (keynote), Professor David Lyon (invited speaker), and Mr Mark Lyell (plenary speaker) at the International Workshop on Sousveillance.



In addressing the audience with opening remarks on the workshop's conception, I began with defining sousveillance and then went on to demonstrate its use. I could think of no better example of sousveillance-at-work than to show a short five minute clip taken by Mann himself in Downtown Toronto (Figure 23.10). In this clip you will note that Steve is exercising his civil rights and pointing out to the police officer on duty that there is a risk of someone getting electrocuted because cables are exposed to pedestrians on the sidewalk. The officer on duty rejects being a subject of Mann's visual recording. He stops Steve as he is nearing him and exclaims: "Sir, you cannot take a picture!" To this Steve questions: "Oh. Why not?" Again, the officer exhorts Steve to stop recording. To this Steve replies- “Ok, I photograph my whole life, I always have…” To this the officer says: “I don’t want to be a part of your life through a photograph. Can you erase that photo please?” Steve does not have a chance to reply at this point and again the officer interjects growing in impatience: “Did you take a picture of me?” Steve replies: “I record my life.” Again the officer extorts: “Did you take a picture of me?” To this Steve makes a correction: “I’m recording video.” The officer interjects several times: “It’s a simple question, did you take a picture of me? Answer the question, yes or no.” Steve admits to taking footage and the officer replies: “Okay, I need you to erase that.” Steve provocatively then says: “Okay, I’ll need to call my lawyer then…” The officer is disgruntled at this point and tells Steve to call his lawyer and to give him his number. The officer continues by insisting: “Do you understand the ramifications of what is going to happen here? Don’t you realise what can happen here?” Steve tells the officer to fill out an incident report about what happened.


Figure 23.10: Code violation and physical assault. Full video at http://wearcam.org/password-66-450.htm.

A struggle over the camera occurs for a good thirty seconds with Steve refusing to let go of it… Steve reassures the officer that he is just recording his whole life. To this the officer replies: “I don’t care about your entire life.” At this point Steve wishes to seek assistance from another officer but he does not wish to do so by leaving his camera behind. Some thirty seconds later after trying to reason with Steve the officer says: “You don’t get it do you… you just don’t get it… What you have on this thing here is *me*, do you understand that?… Doing this job, I don’t want other people to know what I look like and what I do. Now, I don’t know where these pictures are going. They can get on the Internet, they can get somewhere like that and someone can recognise me. And I’ve got a family, and so then my family is in jeopardy. That’s what I’m telling you.” At this point of the exchange, Steve points to the film set camera to try to explain why he is recording. The officer again retorts: “I don’t care about that camera, I care about this thing pointed at my face.” Steve continues to try to convince the officer about the danger of the film set wiring and that he needs the recording for evidence. He suggests going to the police station with the officer but still this does not appease the officer who by this stage is bewildered: “You just don’t get it do you, you don’t, because if you did there wouldn’t be a problem.” To this Steve replies, “I get it… I’ve been recording my life for 20 years.” Again, the officer is adamant: “I’m not a part of your life.” Steve replies, “Well everything that I pass is a part of my life.” The struggle continues and Steve is asked if he has been in trouble with the police before and whether or not he is on any medication. There is no resolution. Steve is hurt in the incident. One week prior to this he had been electrocuted by an above-ground cable exposure in a similar context while a crew was filming at as nearby locale.

This is not the first time that Mann has been the subject of investigation. On a return flight from the United States to Canada, Mann was required by security guards to turn his machine on and off and put it through the X-ray machine while they tugged on his wires and electrodes. The New York Times reported: “the guards took him to a private room for a strip-search in which, he said, the electrodes were torn from his skin, causing bleeding, and several pieces of equipment were strewn about the room" (Guernsey, 2002). Mann was quoted as saying: "We have to make sure we don't go into a police state where travel becomes impossible for certain individuals." At the time, Mann described suffering as a result from the sudden detachment of technology he had worn for decades.

This encounter, and others like it, cut to the core of the implications of sousveillance but also about its everyday role. When we asked Steve what we would do if participants in our glogger assessment were questioned about why they were recording others without their permission, Steve pointed us to his Request for Deletion (RFD) page (Mann, 2009). This is admittedly only a part of the solution. In the future off-the-shelf products might exist to blank out the images of people and car number plates in everyday films, just like in the realm of Google StreetView but for now this is a real issue, as seen in the encounter with the Toronto police officer. No matter how one looks at it however, the increasing use of in-car video recording cameras by law enforcement, business and civilian vehicles, helmet mounted cameras used by motorbike riders, cyclists and extreme sportsmen, roads and traffic authority cameras, embedded cameras in apparatus (e.g. cricket stumps, tazer guns) and the like, mean that through the adoption of sousveillance techniques, the average citizen can reclaim at least some of that asymmetry they have lost. Steve's RFD approach acknowledges however, that an opt-out approach is much more realistic than an opt-in approach. Expecting everyone in my field of view (FOW) to sign a consent form allowing me to film them because I decide to walk the streets wearing a camera is just impossible. But deleting an image or film based on an individual request can be satisfied although it may not always be practical.

We are certain that as social media platforms like glogger proliferate, many will ask:

- To be let alone;
- In what context the footage being taken will be used;

- How the footage taken will be validated and stored; and
- To whom the footage will belong.

These questions are particularly pertinent in the insurance industry as in-car video recording is now widely commercialised (Figure 23.11). Wearables that record like Helmet Cams are becoming plentiful, widely used in the military, extreme sports, the mining sector, and film industry. And we now even have Taser cams, perhaps predated by the stump cam in the game of cricket over a decade ago. Some of these devices, e.g. audio listening spy glasses, are even marketed to minors through school book clubs (Figure 23.12). This raises some interesting questions about how devices used for sousveillance might be misused contra to law in a given jurisdiction. Offences may in fact be committed based on the current law but the law is not yet being enforced to curb activities related to sousveillance. On the other hand, point of view technologies more broadly may even be misused in a stalking capacity or other voyeuristic manner. See for instance, online games marketed to minors that require a webcam to be switched on for play (Figure 23.13). There are also purported borderline cases where a camera is worn by an individual who decides to take footage in a store owned by another person. While it is not a private setting per se, the store owner may not wish for his/her goods and services to be filmed. What are the rights of individuals in public spaces when it comes to private activities? How do we go about a framework for the analysis of any type of surveillance (Clarke, 2012a)?


FIGURE 23.11: Autovision Mobile Media Van in New South Wales, specialising in in-car vehicle tracking, navigation and recording solutions.

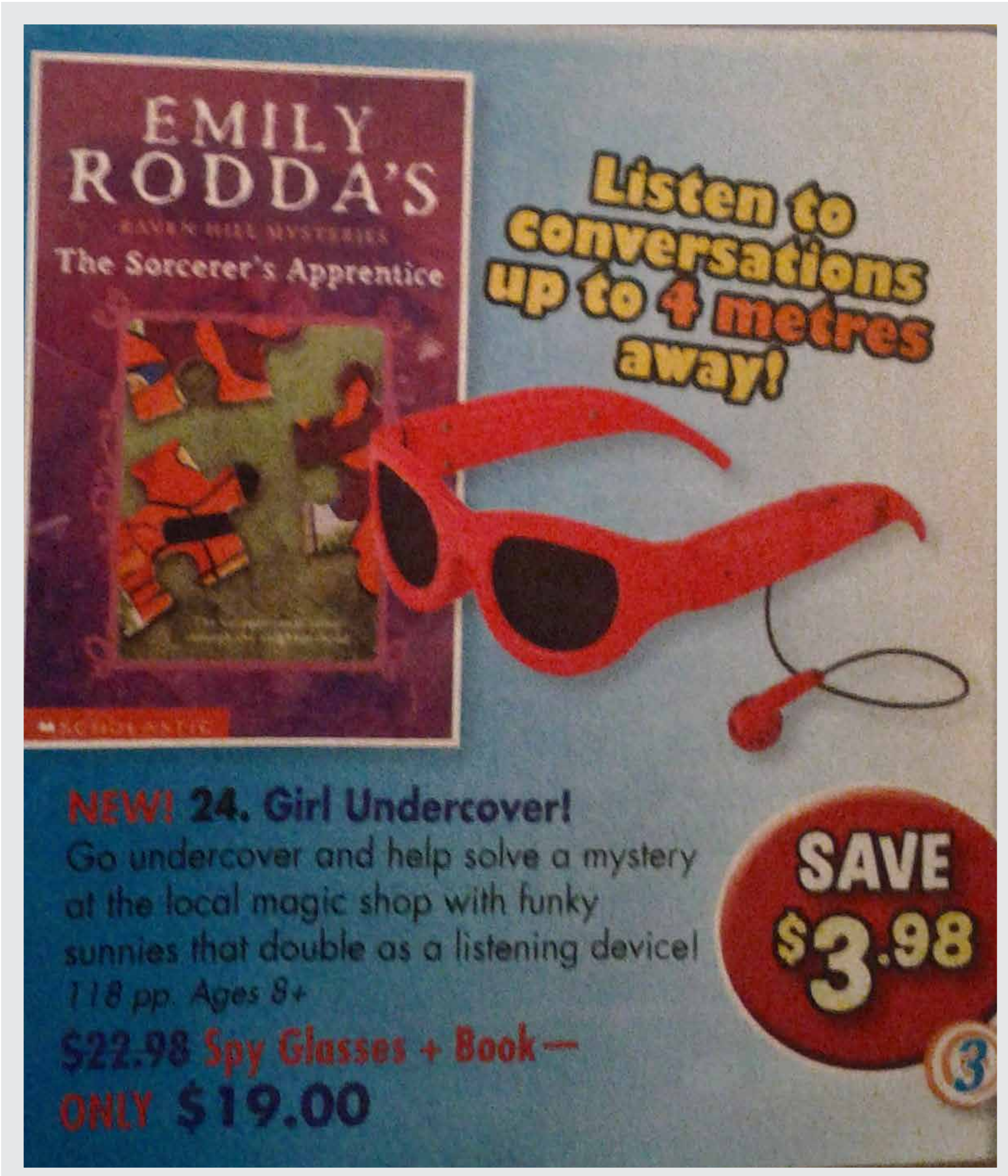



Figure 23.12: Spy glasses plus book marketed to 8+ years olds. Listen to conversations up to 4 metres away.

*Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.*


Figure 23.13: "Angelina Balerina" online dancing game that requires the use of a webcam.

### 23.10.5 Sousveillance and Point of View Technologies in Law Enforcement

The reality is that those supplying point of view (POV) equipment, some specialising in spy equipment, have undergone a massive uptake in demand. This surge is witnessed by the number of organisations that are now dealing in this kind of covert and overt new media and franchising of some of these companies. The statistics indicate that many citizens are now taking matters into their own hands and most probably at the expense of existing legislation to do with Surveillance Devices and Listening Devices Acts. In addition, members of the police force are acquiring their own technology for safety related reasons and incidence/complaint handling in an attempt to reduce the on the job stress they undergo on a daily basis.

In Australia, there were accounts of police officers some 7 years ago, purchasing video camera units and using Velcro to place these cameras in police wagons that had not come equipped with high tech gadgetry to film roadside incidences. But today we are talking about new camera kits that are just not used for in-car recording but for body-worn recording by the police. Police today may wear helmet cams, ear cams, chest cams with audio capability, GPS locators, taser cams etc.

But how much evidence gathering is too much? In the last 12 months we have seen several riots take place -- e.g. the Vancouver Riots and the London Riots. For the first time crowdsourced surveillance played second fiddle to crowdsourced sousveillance. The police called for footage to be submitted for use in convicting rioters for crimes committed. So many thousands of minutes were presented to the police -- above and beyond footage they had taken. One could consider cross-correlation of sorts taking place.

It is predicted however that with time, crowdsourced surveillance will overwhelm the limited resources employed by the police to look at such video evidence -- in many cases potentially thousands of hours worth. Professor Andrew Goldsmith of the Centre for Transnational Crime Prevention at the University of Wollongong, has written about this new visibility with respect to the Tomlinson case. It is the commentators' opinion, as recently recorded in a special guest edited issue of the Journal of Location Based Services (Michael and Michael, 2011), that the police will be moving away from intelligent led policing and toward an IT led policing in near real-time, if not real-time. It will be the ability by the police to say that if you are currently in a zone of public disturbance, or riot, that access to real-time engineering information will be used to denote your location. Additional smartphone modalities will then be harnessed to ascertain whether or not you are a potential perpetrator- for instance accelerometer information that can denote whether you are going up or down stairs or jolting around smashing windows. To borrow from Roger Clarke, this is a form of dataveillance "on the move" (Clarke, 2009).

No doubt this kind of scenario will mean that momentarily people will decide to live off the grid- leave their mobiles behind- or use mobiles on secure and secret platforms like the Blackberry device. But it is exactly this type of scenario that may herald in the age of überveillance- a tiny onboard implant that is injected into the translucent layer of the skin, and records everything as it sees it… implants cannot be left behind… implants are always with you… and implants allegedly do not allow for tampering… the *iplant* as we have termed it, is that 'shock and awe' instrument we have been waiting for to be commercialised in all its spectre. It will supposedly be the answer to all of our electronic health record problems, our social security and tax file numbers, our real-name Internet identity, and secure mobile payments (Michael and Michael, 2012a).

The pitfalls with POV, no matter how many cameras are recording, and no matter from how many perspectives and stakeholders, is that visual evidence has limitations. What is a whole incident? How can we denote past provocation or historical data not available during a given scene? How can we ensure that data on mobile transmission has not been intercepted? How can we ensure data validation? We might well be on another road similar to that of DNA as admissible evidence in a court of law in terms of "eyewitness" recording of events. The key question to ask here is whether or not we can ever achieve "omniscience" through the use of seemingly "omnipresent" new media?

### 23.10.6 References


- Arthur, W. B. (2009). The Nature of Technology: What it is and how it evolves. New York: Free Press.
- Bradwell, J. (2012). Security workshop bring 'sousveillance' under the lens. University of Wollongong: News and Media, http://media.uow.edu.au/news/UOW120478.html?referer=http%3A%2F%2Fworks.bepress.com%2Fkmichael%2F256%2F (6 March 2012).
- Clarke, R. (2009). Roger Clarke's Dataveillance and Information Privacy Home-Page. Rogerclarke.com. Available Online: http://www.rogerclarke.com/DV/ (15 February 2009).
- Clarke, R. (2012a). Point-of-View Surveillance. Rogerclarke.com. Available Online: http://www.rogerclarke.com/DV/PoVS.html (19 March 2012).
- Clarke, R. (2012b). Sousveillance banned in WA. Privacy-List privacy@lists.efa.org.au (18 March 2012).
- Guernsey, L. (2002). At Airport Gate, a Cyborg Unplugged. The New York Times. Available Online: http://www.nytimes.com/2002/03/14/technology/at-airport-gate-a-cyborg-unplugged.html (14 March 2002).
- Leonardo (2009). The Leonardo Awards Program: Award Recipients. Leonardo. Available Online: http://leonardo.info/isast/awardwinners.html (26 August 2009).

- Mann, S. (1995). Wearcam.org as roving reporter; (c) Steve Mann, Feb. 1995. Wearcam.org. Available Online: http://wearcam.org/eastcampusfire.html (17 March 2009).
- Mann, S. (1998). Personal Cybernetics and Intelligent Imaging Systems.Wearcam.org. Available Online http://wearcam.org/ece516/ (18 March 2012).
- Mann, S. (1998). Humanistic Intelligence: WearComp as a new framework for Intelligent Signal Processing. Proceedings of the IEEE. 86 (11) 2123-2151. Available Online: http://hi.eecg.toronto.edu/hi/index.html (8 January 1998).
- Mann, S. (2001). Can Humans Being Clerks make Clerks be Human? - Exploring the Fundamental Difference between UbiComp and WearComp . Informationstechnik und Technische Informatik, 43 (2), 97-106. Available Online: http://wearcam.org/itti/itti.htm (2012).
- Mann, S. (2003). Existential Technology. Leonardo. 36 (1), 19-25. Available Online: http://leonardo.info/isast/articles/Le036-1_mann.pdf (2003).
- Mann, S. (2006). Learning By Being: Thirty Years of Cyborg Existemology. Learning by Being: Thirty Years of Cyborg Existemology. The International Handbook of Virtual Learning Environments. 1571-1592. Available Online: http://wearcam.org/existed/existed.pdf (18 March 2012).
- Mann, S. (2009). Form 698 - Request for Deletion (RFD). Wearcam.org. http://wearcam.org/rfd.htm (19 March 2012).
- Mann, S. (2010). Steve Mann On Cyborg Living and Virtual Reality. Singularity Summit: San Francisco. Available Online: http://www.youtube.com/watch?v=jY2m5Ru1Fm4 (31 August 2010).

- Michael, K. and Michael, M.G. (2009). Innovative Automatic Identification and Location-Based Services: from bar codes to chip implants. New York: Information Science Reference.
- Michael, K. (2011).Event Report: The IEEE Symposium on Technology and Society 2010 (7-10 June): The Social Implications of Emerging Technologies. Journal of Cases on Information Technology. 13 (2), 80-87. Available Online: http://www.igi-global.com/Files/Ancillary/10.4018_jcit.2011040106.pdf (2011).
- Michael, K. and Michael, M.G. (2011). The social and behavioural implications of location-based services. Journal of Location Based Services. 5 (3-4), 121-137. Available Online: http://www.tandfonline.com/doi/pdf/10.1080/17489725.2011.642820 (18 March 2012).
- Michael, K. and Michael, M.G. (2012a). Implementing Namebers Using Microchip Implants: The Black Box Beneath The Skin. In J. Pitt (Ed.) *This Pervasive Day: The Potential and Perils of Pervasive Computing*. London, United Kingdom: Imperial College Press, 101-142.
- Michael, K. and Michael, M.G. (2012b). Program: Sousveillance and the Social Implications of Point of View Technologies in the Law Enforcement Sector. Sixth Workshop on the Social Implications of National Security. Sydney, NSW, Australia. Available Online: http://works.bepress.com/cgi/viewcontent.cgi?article=1290&context=kmichael (22 May 2012).
- PCMag (2009). Definition of Glogger. PCMag.com. Available Online: http://www.pcmag.com/encyclopedia_term/0,2542,t=glogger&i=59866,00.asp (2012).

- Quillian, K. (2004). 2004 Leonardo Award for Excellence Given to Steve Mann. Rhizome. Available Online: http://rhizome.org/discuss/view/15557/ (7 December 2004).
- Shachtman, N. (2003). A Spy Machine of DARPA's Dreams. Wired: Tech Biz|Media. Available Online: http://www.wired.com/techbiz/media/news/2003/05/58909?currentPage=all (20 May 2003).
- Spence, R. (2008). Tiny Camera oh my! Eyeborg. http://eyeborg.blogspot.com.au/2008/11/tiny-camera-oh-my.html (25 November 2008).
- UOW (2010). ISTAS 2010. University of Wollongong. Available Online: http://www.uow.edu.au/conferences/2010/ISTAS/index.html (20 January 2011).
- WA (1998). Surveillance Devices Act 1998 — Sect 6 Available Online: http://www.austlii.edu.au/au/legis/wa/consol_act/sda1998210/s6.html (18 March 2012).

## 23.11 COMMENTARY BY DOUGLAS L. BALDWIN

How to cite this commentary in your report

### Douglas L. Baldwin

*© Douglas L. Baldwin*

In December, 2012, I created a blog that I call "Bugs, Blindness and the Pursuit of Happiness." The blog was set up so that I could interview presenters at the IEEE conference, as well as begin a dialogue concerning wearable computing devices for children in special education. The initial series of blog posts reflects my first interview--a three-day series of discussions with my frie...

Douglas L. Baldwin
Douglas L. Baldwin is a member of The Interaction Design Foundation

Steve Mann provided the vision for a new kind of digital professional; a practitioner capable of prescribing alternative perception. Like any doctor, these specialists would use diagnostic tools to arrive at a diagnosis, from which tailored prescriptions would arise.

After working for over thirty years in special education, as the founder of a vision clinic for children with special needs, and as the founder of a non-profit institute that brought sophisticated navigational technologies to blind and visually impaired kids, it is clear to me that digital perception is a revolution waiting to happen in rehabilitation and special education.

In 2010, I was approached by the X-Prize Foundation to contribute to a proposal concerning portable devices that would benefit the blind, incorporating high technologies into i-pads, cell phones; handheld systems. I knew that these handheld tools were already on the market, or soon to be. I also knew that they would be short of the vision that Steve Mann laid down years ago. There would be no concept of humanistic intelligence, no Eyetap putting sensory input where the brain was expecting to receive it (on the face at eye level), and no attention would be focused toward special education where tailored solutions were the only answer to the needs of unique children.

The list of potential remediations available to the digital perception specialist is extraordinary. I will list five that were outlined for the X Prize-Foundation report. Digital perception specialists will work with the whole body, but my focus is on the potential of what Steve calls Electric Glasses; Eyetap technology that can place computer altered images on the retina in real time.

1. The brain can be directly impacted by electromagnetic variables that alter perception and consciousness. Hemi-synchronization (a sound wave technology) combined with light wave therapies (blue spectrum for wakefulness, and to counter seasonal affective disorder), for example, has the potential to alter mood, affect energy, and assist the evolution of consciousness for individuals.

2. Disabilities, like autism and visual impairment, could be affected (theoretically) by placing laser input on the retina that is augmented, diminished, and/or mediated. For example, stimulating central while in-

hibiting peripheral vision (and the opposite), or emphasizing left field stimulation while inhibiting right field (or reverse) could significantly alter behavior patterns. Cashing input and then slowing it down or speeding it up is a method for determining how altering frame rate affects individual processing and memory storage (i.e. behavior). Enriching or reducing input directly to visual quadrants, hemi-fields, or the whole retina could benefit many people with processing and sensory disabilities.

The entire field of optics will eventually give way to digital image capture and realtime alteration of images reaching the retina. Steve recognizes this when he speaks of downloading visual prescriptions. This switch to digital diagnosis and remediation is a revolution for eye doctors, and eventually portends the demise of the optical industry. Prescriptions will be altered to fit environmental demands in real time, as the environment changes.

The visually impaired and blind populations are handicapped not only by sensory loss, but also by a stark, silent environment. Daniel Kish, CEO of World Access for the Blind, says that blind individuals could navigate the environment as fluidly as the sighted if there was adequate signage available. Kish says that the world is designed to help the sighted navigate. If the environment was smarter (as it could be) and if this smart environment was networked with Steve Mann's Smart Eyeglasses, the blind could navigate without assistance.

Facebook is the beginning of hive brain. Eventually, with molecular implants, hive brain (social networking) will become a reality. In the meantime, Steve Mann's Electric Glasses are an intermediate step, where social networking is at face level. You could look out the eyes of your friends; in a way, become them as their sounds and sights are beamed directly into your perception.

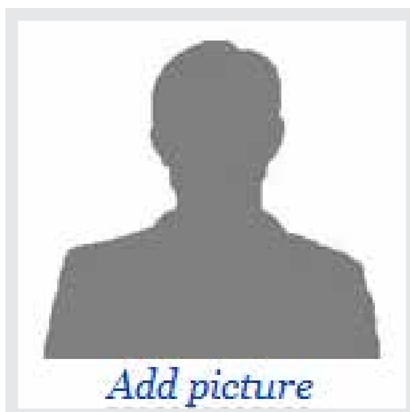
These five examples are entirely possible now. The revolution is overdue. Google's internet glasses may be the door that brings this one step closer to reality. The military has long been working on land warrior goggles. Both Google and the military (and others) have been secret for a long time as these technologies evolved. The Lion is about to be let out of the cage.

My concern remains: Who will transform these technologies into tools that will revolutionize special education and rehabilitation?

## 23.12 COMMENTARY BY WOODROW BARFIELD AND JESSICA BARFIELD

How to cite this commentary in your report

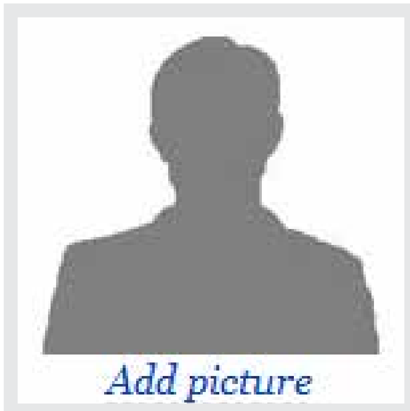
### Woodrow Barfield




Woodrow Barfield served as Professor and Director of the Sensory Engineering Laboratory at the University of Washington. He has degrees in engineering and intellectual property law and has served on the editorial board of Presence and the Virtual Reality Journal....

Woodrow Barfield
Woodrow Barfield is a member of The Interaction Design Foundation

## Jessica Barfield




Jess Barfield is a student and ceramics artist and will be attending Dartmouth College in the fall of 2012. She can normally be found on a field hockey field or in front of a computer....

Jessica Barfield
Jessica Barfield is a member of The Interaction Design Foundation

Steve Mann has written a comprehensive and informative chapter on the general topic of wearable computing (which Steve describes as miniature body-borne computational and sensory devices). We use the phrase- "general topic" because Steve expands his discussion of wearable computing to include the more expansive term, "bearable" computing (essentially wearable computing technology that is *on* or *in* the body). In the chapter, Steve also discusses how wearable computers may be used to augment, mediate, or diminish reality. As background for this commentary, I first met Steve many years ago when I attended a meeting at MIT concerning the first conference to be held on wearable computers, and Steve was then a PhD student at the MIT Media Laboratory (At the conference I made the statement: "Are we wearing the computers, or are they wearing us!"). As the faculty gathered to discuss the aims and direction of the conference, I thought then that Steve had done more to develop the field of wearable computers than the faculty that had gathered to organize the conference. Since my first meeting with Steve, he has continued his innovative work on wearable computing, and he has published extensively on the subject. I particularly enjoyed reading Steve's antidotes concerning his experiences as a "cyborg" in a book Steve wrote for the general public, "*Cyborg: Digital Destiny and Human Possibility in the Age of the Wearable Computer*," 2001. While much of Steve's current chapter is historical in content, he also discusses many of the wearable computing applications he has created, often with Steve's insight as to the rationale behind his inventions.

When we think of the different types of computing technology that may be worn on or in the body, we can envision a continuum that starts with the most basic of wearable computing technology (Steve mentions a wearable abacus) and ends with wearable computing that is actually connected to a person's central nervous system. In fact, as humans are becoming more-and-more equipped with wearable (and bearable) computing technology, the distinction as to what is thought of as a "prosthesis" is becoming blurred as we integrate more computing into human anatomy and physiology. On this very topic, I co-authored a chapter about the use of computing technology to control feedback systems in human physiology ("*Computing Under the Skin*" which was published in Barfield and Caudell, "*Fundamentals of Wearable Computing and Augmented Reality,*" 2001). I agree with Steve that the extension of computing integrated into a person's brain could radically enhance human sensory and cognitive capabilities and alter the direction of human evolution; in fact, in my view, we are just now at the cusp of this development and experimental systems (computing technology integrated into a person's brain) are in-field now that are helping those with severe physical disabilities. For example, consider people with debilitating diseases such that they are essentially "locked in" their own body. With the appropriate wearable computing technology consisting of a microchip that is implanted onto the surface of the brain (where it monitors electronic 'thought' pulses), such people may use a computer by thought alone allowing them to communicate with their family, caregivers, and through the internet, the world at large. Sadly, in the U.S. alone about 5,000 people yearly are diagnosed with just such a disease that ultimately shuts down the motor control capabilities of their body- Amyotrophic lateral sclerosis, sometimes called Lou Gehrig's disease. This disease is a rapidly progressive, invariably fatal neurological disease that attacks the nerve cells responsible for controlling voluntary muscles. Much of the work on control theory and supervisory control of remote robots, along with digital technology, is applicable to the design and use of wearable computing for such individuals.

In our view, anyone at the cutting-edge of their discipline is not only pushing their field further, but by nature of their work, is also at the forefront of other academic disciplines as well. For example, particle physicists in search of the ultimate building blocks of the Universe, often find themselves debating those who hold a nonsecular view of the origins and structure of the Universe. Similarly, Steve's work, albeit on a less dramatic fashion, has raised many important issues of public policy and law. For example, Steve presents the idea that wearable computers can be used to film newsworthy events as they happen or people of authority as they perform their duties. This brings up the question of whether a person has a legal right to film other people in public (answer: generally they do). In the chapter, Steve refers to an interesting case on just this topic decided by the U.S. First Circuit Court of Appeals. In the case, Simon Glik was arrested for using his cell phone's digital video camera (a wearable computer) to film several police officers arresting a young man on the Boston Common. The charges against Glik, which included violation of Massachusetts's wiretap statute and two other state-law offenses, were subsequently judged baseless and were dismissed. Glik then brought suit under a U.S. Federal Statute (42 U.S.C. § 1983), claiming that his very arrest for filming the officers constituted a violation of his rights under the First (free speech) and Fourth (unlawful arrest) Amendments to the U.S. Constitution. The court held that based on the facts alleged, that Glik was exercising clearly-established First Amendment rights in filming the officers in a public space, and that his clearly-established Fourth Amendment rights were violated by his arrest without probable cause. However, the readers of this comment should know: In the U.S. the right to film is not without limitations. It may be subject to reasonable time, place, and manner restrictions a topic in which much case law has been decided.

Steve also discusses privacy issues they may occur when an individual wearing a computer/camera films and records people in public places. While Steve emphasizes the example where state actors, or people generally in positions of authority, are filmed, we worry about the potential to abuse people's privacy using the technology of wearable computing. For example, video voyeurism, the act of filming or disseminating images of a person's "private areas" under circumstance in which the person had a reasonable expectation of privacy regardless of whether the person is in a private or public location, is possible using the technology of wearable computers. In the U.S. such conduct is prohibited under State and Federal law (see for example, *Video Voyeurism Prevention Act of 2004*, 18 U.S.C.A. § 1801). And what about the privacy issues associated with other wearable computing technology such as the ability to recognize a person's face, then search the internet for personal information about the individual (e.g., police record, or credit report), and "tack" that information on the person as they move through the environment? Could digital "scarlet letters" be far off?

Steve's concept of "diminished reality" in which a wearable computer can be used to replace or remove clutter, say for example, an unwanted advertisement on the side of a building, is also of interest to those in law and public policy. On this topic, I published an article in the *UCLA Entertainment Law Review*, 2006, titled-*Commercial Speech, Intellectual Property Rights, and Advertising Using Virtual Images Inserted in TV, Film, and the Real World*. In the article, I discussed the legal ramifications of placing ads consisting of virtual images projected in the real world. We can think of virtual advertising as a form of digital technology that allows advertisers to insert computer-generated brand names, logos, or animated images into television programs or movies; or with Steve's wearable computer technology, the real world. In the case of TV, a reported benefit of virtual advertising is that it allows the action on the screen to continue while displaying an ad viewable only by the home audience. What may be worrisome about the use of virtual images to replace portions of the real world is that corporations and government officials may be able to alter what people see based on political or economic considerations; an altered reality may then become the accepted norm, the consequences of which seem to bring up the dystopian society described in Huxley's "*Brave New World*."

As a final comment, one often hears people discuss the need for "theory" to provide an intellectual framework for the work done in virtual and augmented reality. When I was on the faculty at the University of Washington, my students and I built a head tracked augmented reality system that as one looked around the space of the laboratory, they saw a corresponding computer-generated image that was rendered such that it occluded real objects in that space. We noticed that some attributes of the virtual images allowed the person to more easily view the virtual object and real world in a seamless manner. Later, I became interested in the topic of how people performed cognitive operations on computer-generated images. With Jim Foley, now at Georgia Tech, I performed experiments to determine how people mentally rotated images rendered with different lighting models. This led to thinking about how virtual images could be seamlessly integrated into the real world. I asked the question of whether there was any theory to explain how different characteristics of virtual images combined to form a "seamless whole" with the environment they were projected into, or whether virtual images projected in the real world appeared separate from the surrounding space (floating and disembodied from the real world scene). I recalled a paper I had read while in college by

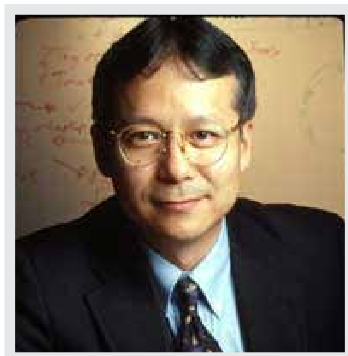
Garner and Felfoldy, published in *Cognitive Psychology*, 1970, on the integrality of stimulus dimensions in various types of information processing. The authors of the paper noted that “separable” dimensions remain psychologically distinct when in combination; an example being forms varying in shape and color. A vast amount of converging evidence suggests that people are highly efficient at selectively attending to separable dimensions. By contrast, “integral” dimensions combine into relatively unanalyzable, unitary wholes; an example being colors varying in hue, brightness and saturation. Although people can selectively attend to integral dimensions to some degree, the process is far less efficient than occurs for separable-dimension stimuli (see also Shepard, R. N., *Attention and the metric structure of the stimulus space, Journal of Mathematical Psychology*, 1964). I think that much can be done to develop a theory of augmented, mediated, or diminished reality using the approach discussed by Garner and Felfody, and Shepard, and I encourage readers of this comment to do so. Such research would have to expand the past work which was done on single images, to virtual images projected into the real world.

Returning to Steve’s chapter, it is an excellent source for those interested in learning about the historical context of wearable computing, and about the numerous applications Steve has developed to design a world in which the humans signal processing capabilities and wearable computing system functions form a feedback loop; the thought being, two brains are better than one! We also see Steve’s work evolving in the not too distant future to the point where humans and wearable computing technology “live” in a mutually symbiotic manner, which implies of course, the primary thinker, the wearable computing, is in some way benefiting from having a human in the loop. So, returning to what I said at the first conference held on wearable computers: Are we wearing them, or are they wearing us?

## 23.13 COMMENTARY BY HIROSHI ISHII

How to cite this commentary in your report

### Hiroshi Ishii



Hiroshi Ishii is Jerome B. Wiesner Professor of Media Arts and Sciences at the MIT Media Lab, where he is head of the Tangible Media group and co-director of the Things That Think (TTT) consortium. Ishii's research focuses upon the design of seamless interfaces between humans, digital information, and the physical environment. His group seeks to change the "painted bits" of GUIs to "tang...

Hiroshi Ishii
Hiroshi Ishii is a member of The Interaction Design Foundation

Although most people now see Steve Mann as a father of Wearable Computing, he was already far beyond "wearable" some 20 years ago when I first met this cyborg in the MIT Media Lab. His series of inventions was not just about "wearable," but a radical form of symbiosis and co-evolution of machine and human being, which he has been experimenting with for more than two decades, living in symbiosis with computations on his skin, eyes, ears, and in his soul. His vision of "Mediated Reality" has the same significance as the "Collective Intelligence" vision of Douglas Engelbart. Because of the same reason that thinking of Doug as an inventor of the Mouse is inappropriate and disrespectful, seeing Steve as merely an inventor of "wearable" is not right. Please read his visionary papers, and enjoy Steve Mann's deep universe.

## 23.14 REFERENCES

*Bade, J. Peter, Jr., Gerald Quentin Maguire and Bantz, David F. (1990). The IBM/Columbia Student Electronic Notebook Project. IBM, T. J. Watson Research Lab, Yorktown Heights, NY, USA*

*Barfield, Woodrow and Caudell, Thomas (eds.) (2001): Fundamentals of Wearable Computers and Augmented Reality. CRC Press*

*Bass, Thomas A. (2000): The Eudaemonic Pie. iUniverse*

*Brin, David (2011). Sousveillance: A New Era for Police Accountability. Retrieved 25 December 2011 from http://davidbrin.blogspot.com/2011/06/sousveillance-new-era-for-police.html*

*Dennis, Kingsley (2008): Keeping a Close Watch - The Rise of SelfSurveillance & the Threat of Digital Exposure. In Sociological Review, 56*

*Feiner, Steven K., MacIntyre, Blair and Seligmann, Dorée D. (1993): Knowledge-Based Augmented Reality. In Communications of the ACM, 36 (7) pp. 53-62*

*Geary, James (2002): The Body Electric: An Anatomy of the New Bionic Senses. Rutgers University Press*

*Herling, Jan and Broll, Wolfgang (2011). Video: Page title suppressed - page not yet published. Retrieved 6 June 2013 from [URL suppressed - page not yet published]*

*Knight, Brooke A. (2000): Watch Me! Webcams and the Public Exposure of Private Lives. In Art Journal, 59 (4)*

*Mann, Steve (2011). Video: Page title suppressed - page not yet published. Retrieved 6 June 2013 from [URL suppressed - page not yet published]*

*Mann, Steve (2003): Existential Technology: Wearable Computing is not the real issue. In Leonardo, 36 (1) pp. 19-25*

*Mann, Steve (2004): "Sousveillance": inverse surveillance in multimedia imaging. In: Schulzrinne, Henning, Dimitrova, Nevenka, Sasse, Martina Angela, Moon, Sue B. and Lienhart,* *Rainer (eds.) Proceedings of the 12th ACM International Conference on Multimedia October 10-16, 2004, New York, NY, USA. pp. 620-627*

Mann, Steven (1998): *Humanistic computing: "WearComp" as a new framework and application for intelligent signal processing. In Proceedings of the IEEE, 86 (11) pp. 2123-2151*

Mann, Steve (1996b). *Wearable, tetherless computer--mediated reality: WearCam as a wearable face--recognizer, and other applications for the disabled, MIT Tech Report number 361, Cambridge, Massachusetts. MIT http://wearcam.org/vmp.htm*

Mann, Steve (1996a): *Smart Clothing: The Shift to Wearable Computing. In Communications of the ACM, 39 (8) pp. 23-24*

Mann, Steve (2001b): *Intelligent Image Processing. Wiley-IEEE Press*

Mann, Steve (2001a): *Guest Editor's Introduction: Wearable Computing-Toward Humanistic Intelligence. In IEEE Intelligent Systems, 16 (3) pp. 10-15*

Mann, Steve and Niedzviecki, Hal (2001): *Cyborg: Digital Destiny and Human Possibility in the Age of the Wearable Computer. Doubleday of Canada*

Mann, Steve, Nolan, Jason and Wellman, Barry (2003): *Sousveillance: Inventing and Using Wearable Computing Devices for Data Collection in Surveillance Environments. In Surveillance & Society, 1 (3)*

Mulligan, Deirdre K. (2009). *University Course on Sousveillance: SURVEILLANCE, SOUSVEILLANCE, COVEILLANCE, AND DATAVEILLANCE. Retrieved 1 December 2011 from Berkeley iSchool:*

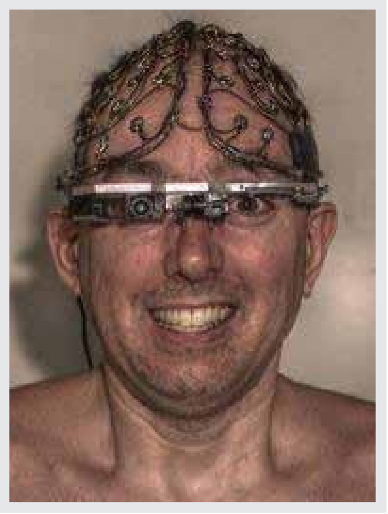
Roberts, Steven K. (1988): *Computing Across America: The Bicycle Odyssey of a High-Tech Nomad. Information Today Inc*

Thompson, Clive (2011). *Clive Thompson on Establishing Rules in the Videocam Age. Retrieved 1 December 2011 from Wired Magazine: http://www.wired.com/magazine/2011/06/st_thompson_videomonitoaring/*

Thorp, Edward O. (1966): *Beat the Dealer: A Winning Strategy for the Game of Twenty-One. Vintage*

## About the author

### Steve Mann



*© Steve Mann*

Personal Homepage:
http://www.eecg.utoronto.ca/~mann/

Steven Mann is a tenured professor at the Department of Electrical and Computer Engineering at the University of Toronto. Mann holds degrees from the Massachusetts Institute of Technology (PhD in Media Arts and Sciences '97) and McMaster University, where he was also inducted into the McMaster University Alumni Hall of Fame, Alumni Gallery, 2004, in recognition of his career as an inventor and teacher. While at MIT he was one of the founding members of the Wearable Computers group in the Media Lab. In 2004 he was named the recipient of the 2004 Leonardo Award for Excellence for his article "Existential Technology," published in Leonardo 36:1.

## Your notes and thoughts on chapter 23

Record your notes and thoughts on this chapter. If you want to share these thoughts with others online, go to the bottom of the page at: http://www.interaction-design.org/encyclopedia/wearable_computing.html

### NOTES

# Chapter 24

# Socio-Technical System Design

***by Brian Whitworth with Adnan Ahmad.***

A socio-technical system (STS) is a social system operating on a technical base, e.g. email, chat, bulletin boards, blogs, Wikipedia, E-Bay, Twitter, Facebook and YouTube. Hundreds of millions of people use them every day, but how do they work? More importantly, can they be designed? If socio-technical systems are social and technical, how is computing both at once?

This chapter may be used as part of a STS design course. Hence each part has a set of interesting discussion questions that students can investigate and report back to the class. Anyone wishing to set up a course in the design of social technologies is welcome to use this resource

## 24.1 PART 1: THE EVOLUTION OF COMPUTING

"Evolution is systems evolving higher levels"

### 24.1.1 A short history

The first computer was conceived of as a machine of cogs and gears (Figure 24.1). It became operational in the 1950s and -60s with the invention of semi-conductors. In the 1970s, a *hardware* company called IBM[1] was a computing leader. In the 1980s software became more important, so by the 1990s a *software* company called Microsoft[2] took the computing lead, giving ordinary people tools like word-processing. During the 1990s, computing became more *personal,* as the World-Wide-Web turned Internet URLs into web site names that people could read[3]. Then a company called Google[4] offered the ultimate personal service, free access to the vast public library we call the Internet, and everyone's gateway to the web became the new computing leader. The 2000s computing evolved yet again, to become a *social* medium as well as a personal tool. So now Facebook challenges Google, as Google challenged Microsoft, as Microsoft challenged IBM.

1. IBM stands for International Business Machines
2. Microsoft stands for microcomputer software
3. IP addresses like 208.80.154.225 became Uniform Resource Locator (URL) names like http://en.wikipedia.org/
4. Google stands for a 1 followed by 100 zeros, i.e. a very large number

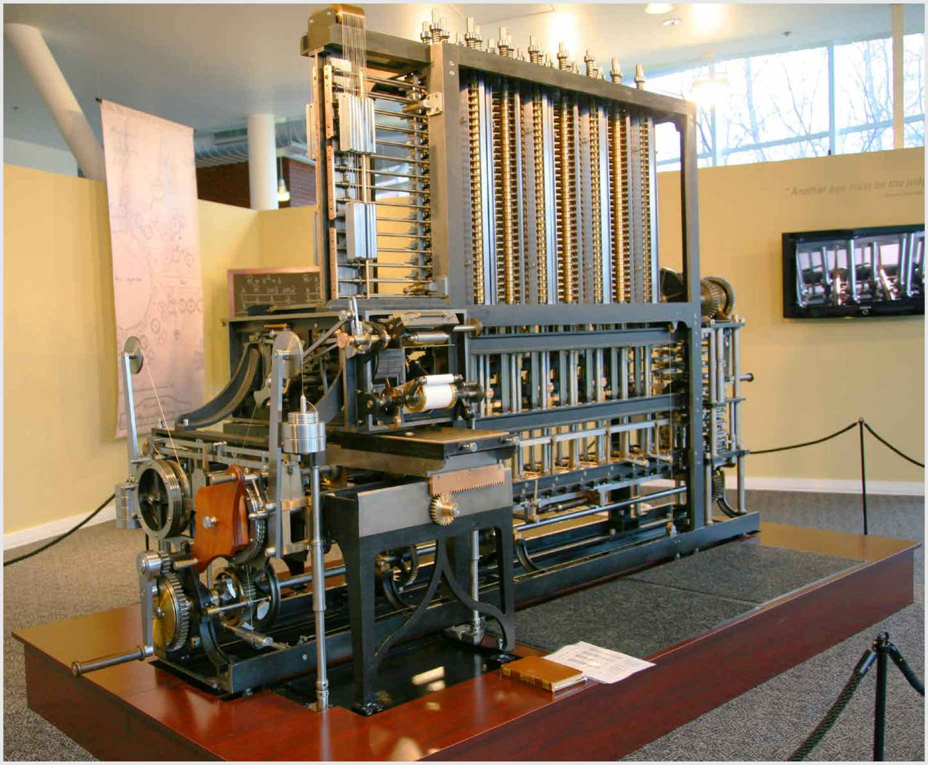
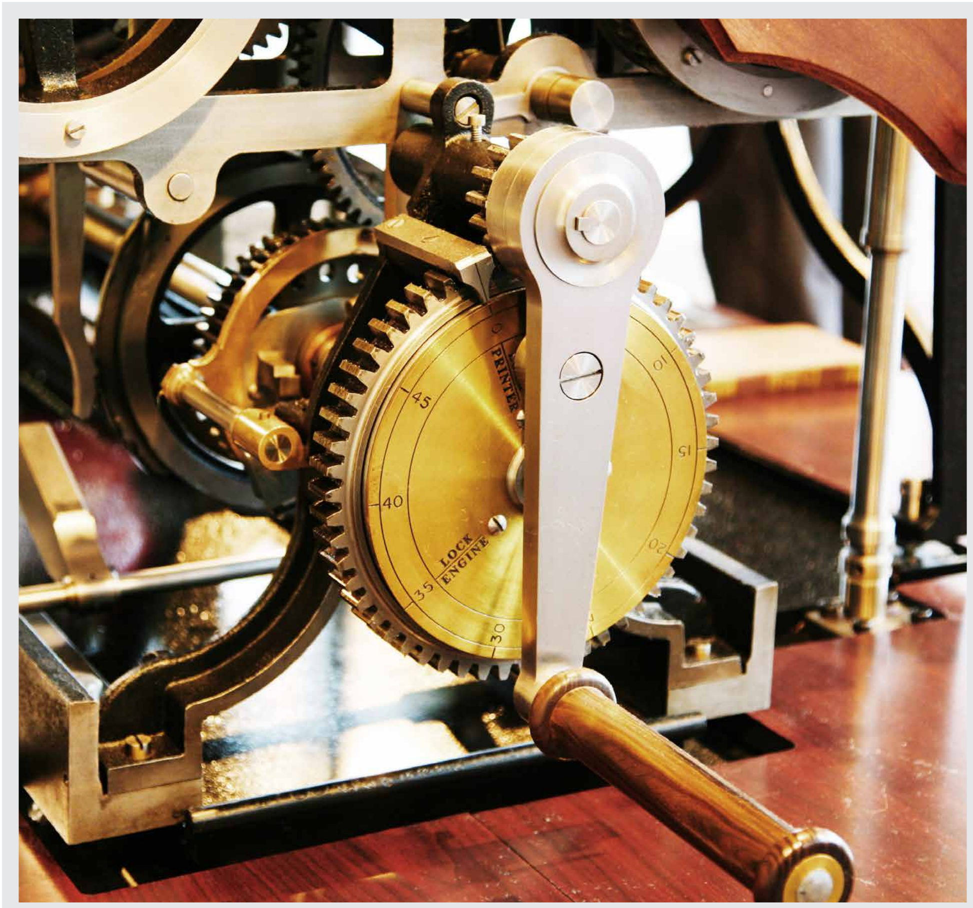


Figure 24.1: Charles Babbage (1791-1871) designed the first automatic computing engines. He invented computers but failed to build them. The first complete Babbage Engine was completed in London in 2002, 153 years after it was designed. Difference Engine No. 2, built faithfully to the original drawings, consists of 8,000 parts, weighs five tons, and measures 11 feet. The one pictured above is Serial Number 2 and is located in Silicon Valley at the Computer History Museum in Mountain View, California.




*Courtesy of Jitze Couperus. Copyright: CC-Att-SA-2 (Creative Commons Attribution-ShareAlike 2.0 Unported).*

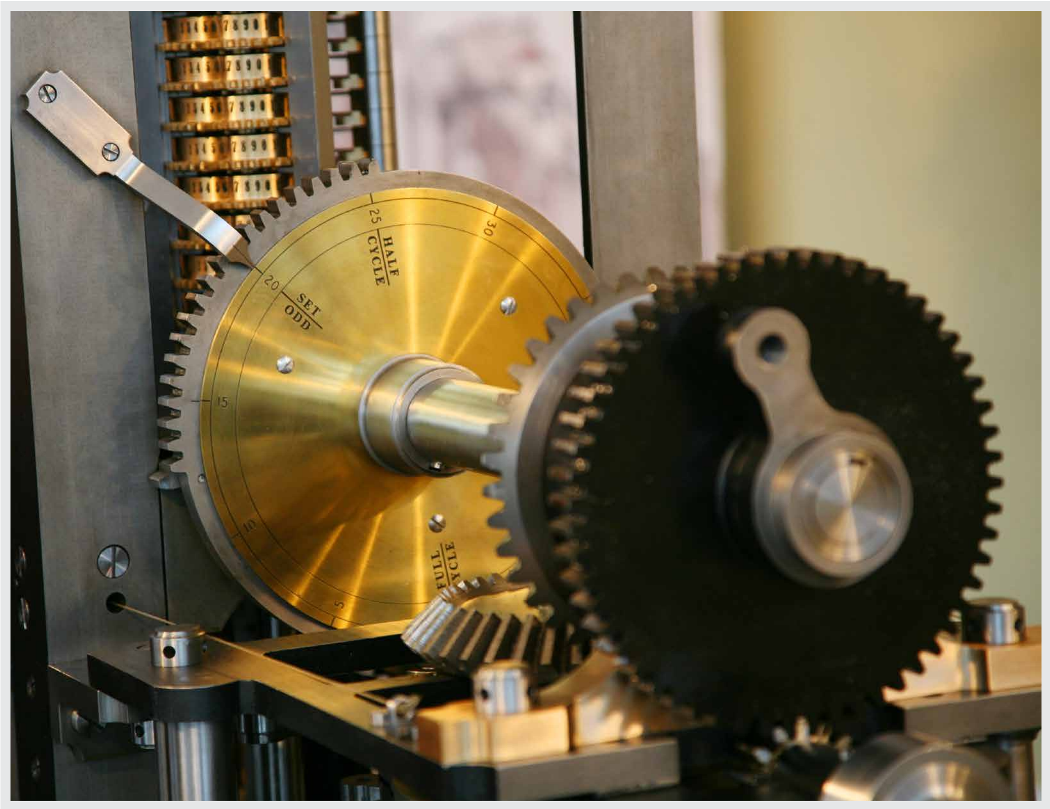





Figure: Details from Babbage's difference engine.

Computing has re-invented itself every decade or so (Figure 24.2). What began as just hardware became about software, then people, and now communities. A physical machine exchanging electricity became software exchanging information, people exchanging meaning and now communities exchanging memes[5]. The World Wide Web was initially an information web (Web 1.0), then an active web (Web 2.0), now a semantic web (Web 3.0) and is becoming a social web (Web 4.0). Each evolutionary step built on the previous, as social computing needs personal computing, personal computing needs software and software needs hardware.

5. A meme is an idea, behavior or style communicated within a culture

The corresponding evolution of computing design culminates in socio-technical design.

When the software era arrived, hardware continued to evolve but hardware leaders like IBM no longer dominated computing as before. The evolution of computing changed business fortunes by changing what computing *is*. Selling software makes more money than selling hardware because it changes more often. Web queries are even more volatile, but Google gave a service away for free and then sold advertising around it — it sold its services to those who sold theirs. The business model changed, because selling knowledge is not like selling software. Facebook's business model is still evolving. It now challenges Google because we *relate* to family and friends more than we *query* knowledge - social exchanges have more trade potential than knowledge exchange.

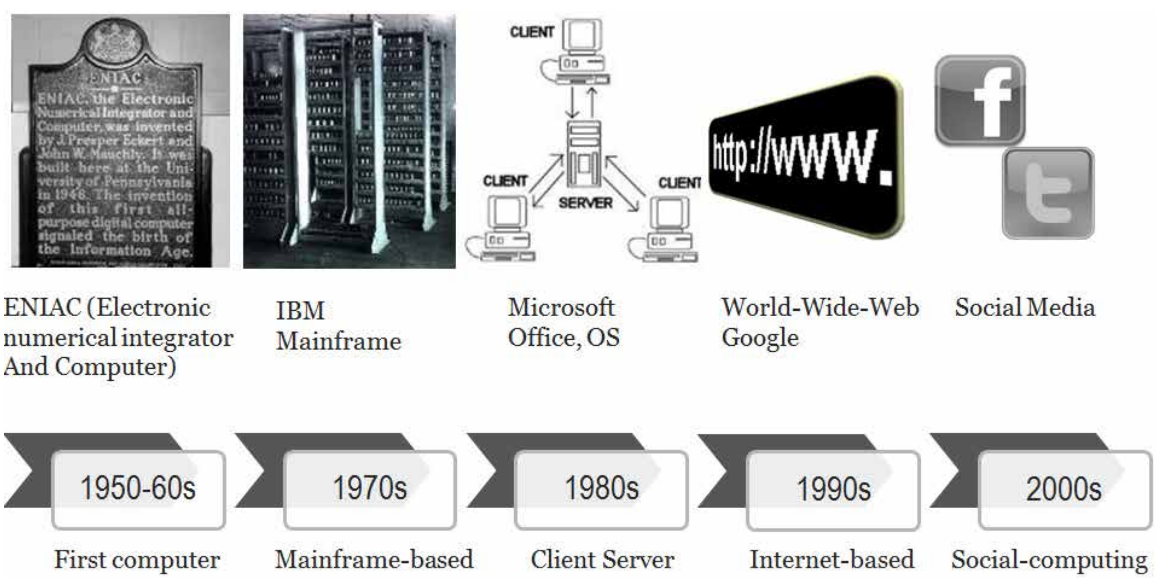



Figure 24.2: The computing evolution.

Yet friends are just social dyads. It is naive to think that social computing will stop at a unit of two. Beyond friends are tribes, cities, city-states, nations and meta-nations like the European Union. A community isn't like a friend, as one *has* a friend but *belongs* to a community. With a world population at seven billion and growing, Facebook's 900 million active accounts are just the beginning. The future is computer support for acting groups, families, tribes, nations and eventually a global community, e.g. a group browser for people to tour the Internet together, commenting to each other as they go. Each could take turns to pick the next site or follow an expert host. If socio-technology is just beginning, we need to understand how it works.

### 24.1.2 Computing levels

The basis of socio-technical design is general systems theory (Bertalanffy, 1968). It describes what the disciplines of science have in common: sociologists see social systems, psychologists cognitive systems, computer scientists information systems and engineers hardware systems. All refer to systems. In general systems theory, no discipline has a monopoly on science and all are valid. Discipline isomorphies[6] arise from common system properties, e.g. a social agreement measure that matched a biological diversity measure (Whitworth, 2006). Mechanical, logical, psychological and social systems are studied by engineers, computer scientists, psychologists and sociologists respectively. These perspectives in computing give levels (Table 24.1). Computing then began at the mechanical level, evolved an information level, then acquired human and community levels.

6. A discipline isomorphy is when different fields with different forms present the same equation or law

| Level | Examples | Discipline |
|---|---|---|
| *Community* | Norms, culture, laws, zeitgeist, sanctions, roles | Sociology |
| *Personal* | Semantics, attitudes, beliefs, feelings, ideas | Psychology |
| *Informa-tional* | Programs, data, bandwidth, memory | Computer science |
| *Mechanical* | Hardware, motherboard, telephone, FAX | Engineering |

Table 24.1: Computing levels as discipline perspectives.

Levels also help clarify terminology. In Figure 24.3, a *technology* is any tool people build to use, e.g. a spear is a technology[7]. So a hardware device alone is a technology, but *information technology* (IT) is both hardware and software. Likewise, *computer science*(CS)[8] is a hybrid of mathematics and engineering, not either alone. So information technology is not a sub-set of technology, nor is computer science a sub-set of engineering.

*Human computer interaction* (HCI) is then a person plus an IT system, with physical, informational and psychological levels. Just as IT isn't hardware, so HCI isn't IT, but the child of IT and psychology. HCI links CS to psychology as CS linked engineering to mathematics. HCI introduces human requirements to computing and HCI systems turn information into meaning.

Finally, people can form an online community with hardware, software, personal and community levels. If the first two levels are technical and the last two so-

7. Anything we use physically is technology, e.g. a table is technology

8. The study of information processing

cial, the result is *a socio-technical system* (STS). If technology design is computing built to hardware and software requirements, then socio-technical design is computing built to personal and community requirements *as well*. In socio-technical systems, the new "user" of computing is the community (Whitworth, 2009b).

Currently, many terms refer to human factors in computing: Engineers extend the term IT to refer to applications built to user requirements; Business calls people and organizations using computing information systems (IS); Education prefers information communication technology(ICT) to describe computer communication; Health chose the term informatics. Whether your preferred term is IT, IS, ICT or informatics doesn't change the basic idea, that people are now part of computing. This chapter uses the term HCI for consistency[9].

In this pan-discipline view, *all* of Figure 24.3 is computing, whose complexity arises from its discipline promiscuity. Socio-technology then designs a computer product as a social *and* technical system. Limiting computing to hardware (engineering) or software (computer science) denies its obvious evolution.

Levels in computing are ways to view it, not ways to partition it, e.g. a pilot in a plane is *one* system with different levels, *not* a mechanical part (the plane) plus a human part (the pilot). The physical level includes not just the plane body but also the pilot's body, as both have weight, volume etc. The information level isn't just the onboard computer, but also the neuronal processing in the pilot's brain that generates the *qualia*[10] of human experience.

9. An alternative term is CHI, or computer-human interaction.

10. A qualia is a basic subjective experience, e.g. the pain of a headache

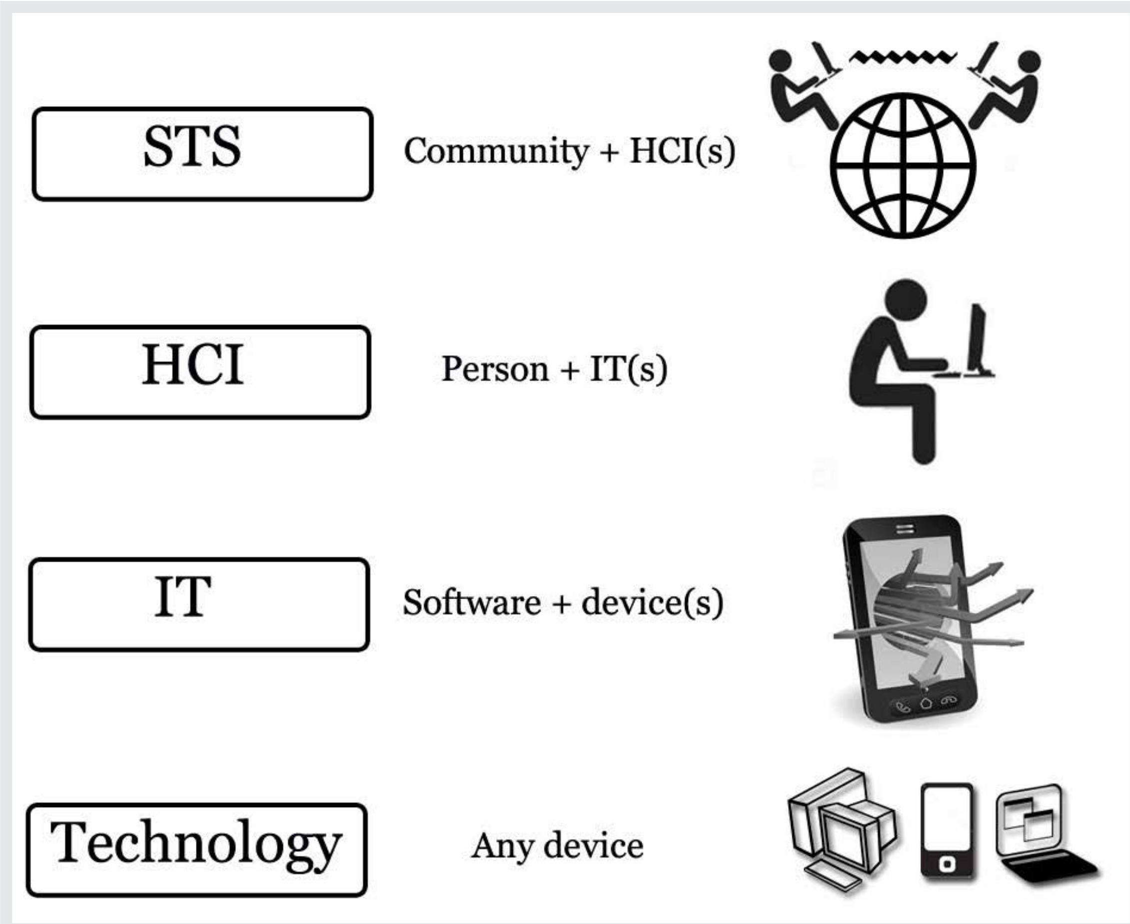



FIGURE 24.3: Computer system levels.



The human level is just a pilot who sees the plane as an extension of his or her body, like extra hands or feet, and computer data like extra eyes or ears. On this level, the pilot is the actor and the plane is just a tool. The information level covers all processing, not just of onboard computers but also of the brain. The physical level is not just the body of the plane but also of the pilot. In an aerial conflict, the tactics of a piloted plane will be different from a computer drone. Finally, a plane in a squadron may do things it would not do alone, e.g. expose itself as a decoy so others can attack the enemy.

### 24.1.3 The reductionist dream

The reductionist dream, based on logical positivism[11], is that only the physical level is "real", so everything else must reduce to it. Yet when Shannon and Weaver defined information as a choice between physical options, the options were physical but the choosing wasn't (Shannon and Weaver, 1949). A message physically fixed in *one* way has by this definition *zero* information, because the other ways it *could* have been don't exist physically[12]. It is strange but logically true that hieroglyphics one can't read have in themselves no information at all. It is reader choices that generate information, which until deciphered is unknown. If this were not so, data compression couldn't put the same data in a physically smaller signal, which it can. Information is defined by the encoding, not the physical message. If the encoding is unknown, the information is undefined, e.g. an electronic pulse sent down a wire could be a bit, or a byte (an ASCII "1"), or as the first word of a dictionary, say Aardvark, be many bytes. The information a message conveys depends on the decoding process, e.g. every 10th letter of a text gives a new message. Information doesn't exist physically, as it can't be touched or seen. Physicality is necessary for it, but not sufficient.

That mathematical laws are real even though they aren't concrete is *mathematical realism* (Penrose, 2005). Mathematics is a science because its constructs are logical, not because they are physical. They are real because we conceive them, not because they physically exist. That they are later physically useful is another matter. *Cognitive realism* is the case that cognitions are also real because we experience them. Mathematical or cognitive constructs defined in physical terms become empirical[13], and so the feedback loop of science still works, e.g. fear mea-

11. Logical positivism is a nineteenth century meta-physical position stating that all science involves only physical observables. In psychology, it led to Behaviorism (Skinner, 1948) which is now largely discredited (Chomsky, 2006). Science is not a way to prove facts, but a way to use world feedback to make best guesses. See researchroadmap.org for more details

12. An on/off line voltage choice is one bit, but a physical 'on' signal alone is no information

13. Empirical means derived from the physical world. Mental constructs with no physical referent, like love, are outside it.

sured by heart rate is a cognitive construct measured in physical terms. Yet fear isn't just heart-rate, as it can also be measured by pupil dilation, blood pressure, etc. Even terms like "red" aren't physical facts as the light spectrum is continuous, with no red frequency section.

The physical level alone is what it is. It has no choices so has no information, i.e. *reductionism denies information science*. In physics, it gave a clockwork universe, where each state perfectly defined the next. Quantum theory flatly denied this, as quantum events are by definition random, i.e. explained by no physical history. Either quantum theory is wrong, which it has never been, or reductionism, that only the physical is real, is a naive nineteenth century assumption that has had its day. If all science were physical, all science would be physics, which it is not.

Physics today has a *quantum* level, i.e. a primordial non-physical[14] reality below physical reality (Whitworth, 2011). Yet long ago, the great 18th Century German philosopher Kant argued that reality is just a view, that we don't see *things as they are in themselves* (Kant, 1999)[15]. Levels return the observer to science, as quantum theory's measurement paradoxes demand. In philosophy, psychology, mathematics, computing and quantum physics, levels apply[16].

---

14. For example, quantum collapse ignores the speed of light limit and quantum waves travel many paths at once

15. He called a 'thing in itself' the noumenon, as opposed to the phenomenon, or the view we see. A bat or a bee would see the world differently from us. It is egocentrism to assume the world is only as we see it.

16. With a non-physical quantum reality below the physical

### 24.1.4 Science as a world view


Figure 24.4: Computing levels as abstract views.



A level is a world view, a way of seeing complete and consistent in itself. In the mechanical view, a computer is *all* hardware, but in the informational view it is *all* data. One can't point to a program on a motherboard nor a device in a data structure. A mobile phone doesn't have hardware and software parts, but is hardware or software *in toto*. Hardware and software are ways to look at it, not ways to divide it up. Hardware becomes software when we view computing in a different way. The switch is as one swaps glasses to see the same object close-up. The disciplines of science are world views, like walking around an object to see it from different perspectives.

Levels are a fact of science, e.g. to describe World War II as a "history" of atomic events would be ridiculous. A political summary is more useful. Yet levels *emerge* from each other, as higher abstractions form from lower ones (Figure 24.4). Information needs hardware choices, cognitions need information flows, and communities need common cognitions. Conversely, without physical choices there is no information, without information there are no cognitions and without cognitions there is no community[17].

A world view has properties, like being:

1. Essential.One cannot view a world without first having a point of view.
2. *Empirical*. Based on world interaction, e.g. information is empirical.
3. *Complete*. A world view consistently describes a whole world.
4. *Subjective*. One *chooses* a view before viewing, explicitly or implicitly.
5. *Exclusive*. One can't view two ways at once, as one can't sit in two places at once.[18]
6. *Emergent*. One world view can emerge from another.

Levels as views must be chosen before viewing, i. e. *pick a level then view*.

Yet how we see the world affects how we act, e.g. if we saw ultra-violet light, as bees do, previously dull flowers would become bright. Every flower shop would have to change its stock. *Levels as higher ways to view a system are also new ways to operate and design it*, e.g. new software protocols like Ethernet can improve network performance as much as new cables.

New ways to view computing affect how we build it, and how social levels affect technology design is socio-technical design. Level requirements cumulate, so socio-technical design includes hardware, software and HCI requirements

17. A community is a set of people who see themselves as a social unit.

18. Or as one can't lever from two fulcrums at once. One can, of course, view from one perspective then another

(Figure 24.5). What *appears* as just hardware now has requirements outside itself, e.g. smart-phone buttons mustn't be too small for people's fingers. Levels are why computer design has evolved from hardware engineering to socio-technology.

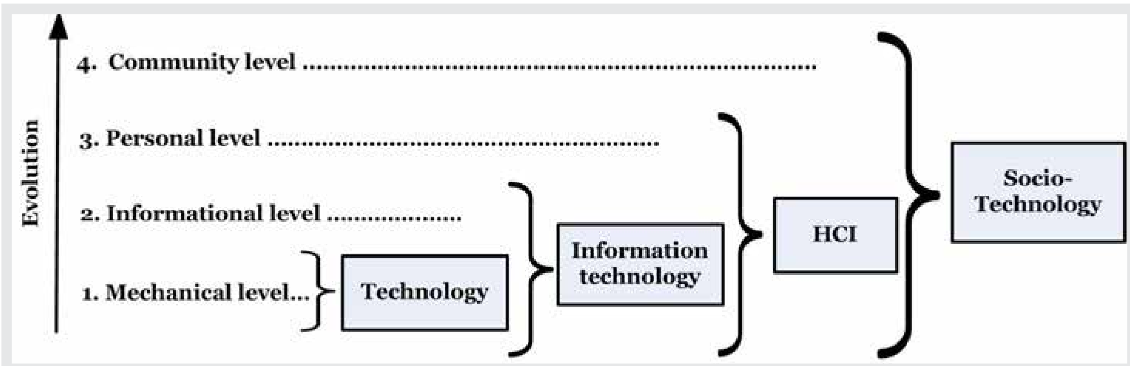



Figure 24.5: Computing applications and levels.



For a village beside a factory, community needs come second to factory productivity, with ethics an after-thought, but for socio-technology, the community and the technology are one. If social needs are not met there is no community, and if there is no community the technology fails to perform as expected. Socio-technical design is the application of community requirements to people, software and hardware. The following sections derive each computing level from the previous.

### 24.1.5 From hardware to software

Hardware is any physical computer part, e.g. mouse, screen or case. It doesn't "cause" software nor is software a hardware output, as physical systems have physical outputs. *We* create software by seeing choice in physicality. Software needs hardware but it isn't hardware, as the same code can run on a PC, Mac or mobile phone. An entity relationship diagram can work for any physical storage, whether disk, CD or USB, as data entities aren't disk sectors. Software assumes some hardware but no specific one.

If any part of a device acquires software, the whole system gets an information level, e.g. a computer is information technology even though its case is just hardware. We describe a system by its highest level, so if the operating system "hangs"[19] we say "the computer" crashed, even though the *computer hardware is working fine*. Rebooting fixes the software problem with no hardware change, so a software system can fail while the hardware still works perfectly.

Conversely, a computer can fail as hardware but not software, if a chip overheats. Replace the hardware part and the computer works with no software change needed. Software can fail without hardware failing and hardware can fail without software failing. New hardware needn't change software and new software needn't change hardware. Each level has its own performance requirements: if software fails we call a programmer, but if hardware fails we call an engineer.

Software *requirements* can be met by hardware *operations*, e.g. reading a logical file takes longer if the file is fragmented, as the drive head must jump between physically distant disk sectors. Defragmenting a disk improves software access by putting files in adjacent physical sectors. File access improves, but the physical drive read rate hasn't changed, i.e. hardware actions can meet software goals, e.g. database and network requirements gave new hardware chip commands. The software goal, of better information throughput, also becomes the hardware goal, e.g. physical chip design today is as much about caching and co-processing as it is about cycle rate.

### 24.1.6 From software to HCI

HCI began with the personal computing era. Adding people to the computing equation meant that getting technology to work was only half the problem - the other half

19. If the software gets in an infinite loop, we say it 'hangs'

was getting people to use it. Web users who didn't like a site just clicked on. Web sites that got more hits succeeded because given equal functionality, users chose the more usable product (Davis, 1989), e.g. Word replaced Word Perfect because it was more usable - users who took a week to learn Word Perfect picked up Word in a day. As computing previously gained a software level, it now gained a human level.

*Human computer interaction* (*HCI*) is a person using IT, as IT is software using hardware. As computer science merges mathematics and engineering, but is neither, so HCI merges psychology and computer science, but is neither. Psychology is the study of people, and computer science the study of software, but the study of people using software, or HCI, is new. It is another computing discipline that cuts across other disciplines. HCI applies psychology to computing design, e.g. Miller's paper on cognitive span suggests limiting computer menu choices to seven (Miller, 1956). Our many senses and multi-media computing is another example of a human requirement defining computing.

### 24.1.7 From HCI to STS

Social structures, roles and rights add a fourth level to computing. Socio-technical design uses the social sciences in computing design as HCI uses psychology. STS is not part of HCI, nor is sociology part of psychology, because a society is more than the people in it, e.g. East and West Germany, with similar people, performed differently as communities, as is true for North and South Korea today. To say "the Jews" survived but "the Romans" didn't is to say that the society didn't continue, not its people, as no Roman era people are alive today. A society is not just the people in it. People who gather to view a spectacle or customers coming to shop for bargains, are not a community. A community is here *an agreed form of social interaction that persists* (Whitworth and de Moor, 2003).

Social interactions can have a physical or a technical base, e.g. a *socio-physical system* is people connecting by physical means. Face-to-face friendships cross seamlessly to Facebook because the social level persists across physical and electronic architecture bases. Whether electronically or physically mediated, a social system is *always* people interacting with people. Electronic communication may be “virtual”, but the people involved are real.

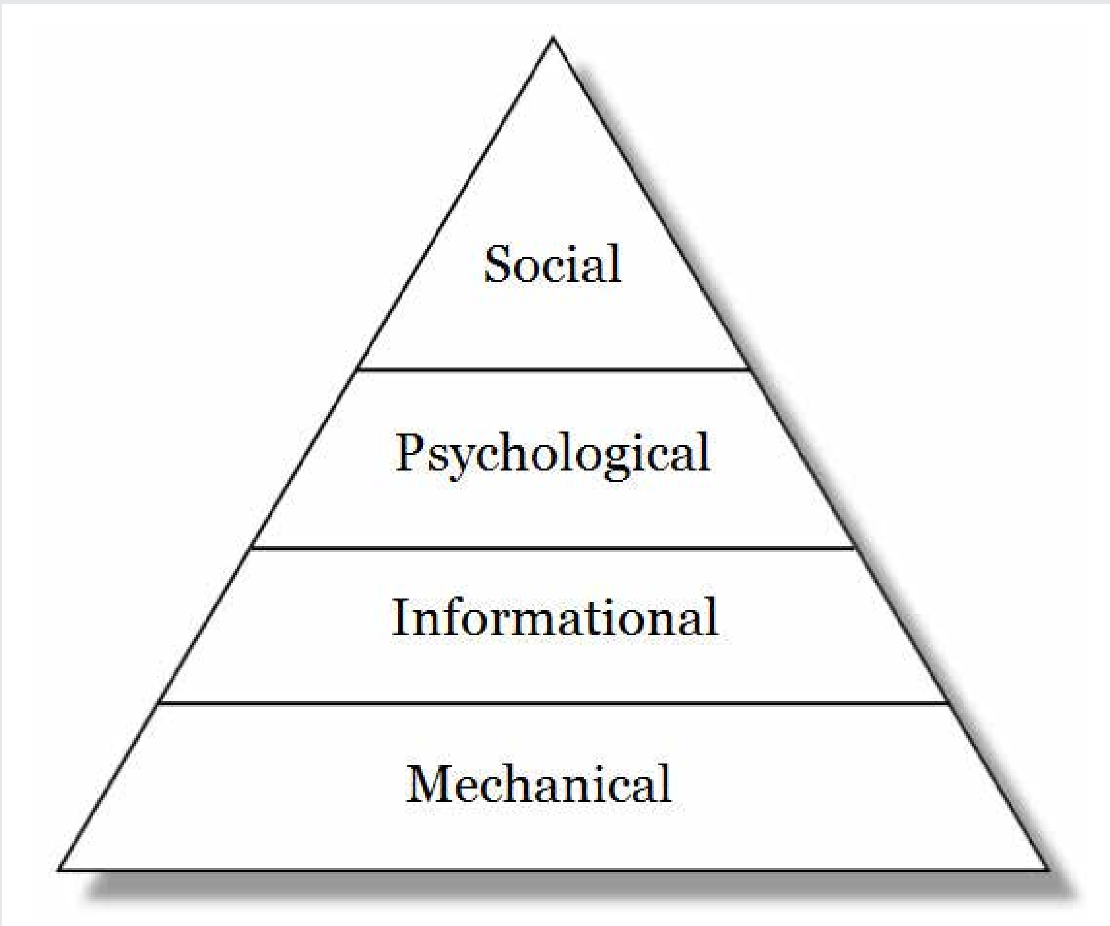



FIGURE 24.6: The computing requirements hierarchy.



A community works through people using technology as people work through software using hardware, so social requirements are now part of computing design (Sanders and McCormick, 1993). While sociology studies the social level alone, socio-technical design studies how personal and social requirements can be met by IT system design. Certainly this raises the cost of development, but then systems like social networks have far more performance potential.

### 24.1.8 The computing requirements hierarchy

The evolution of computing implies a requirements hierarchy (Figure 24.6). If the hardware works software becomes the priority, if the software works user needs arise, and when user needs are met social requirements follow. As one level's issues are met those of the next appear, as climbing one hill reveals another. As hardware over-heating problems are solved, software data locking problems arise. As software response times improve, user response times become the issue. Companies like Google and E-bay still seek customer satisfaction, but customers in crowds have social needs like fairness and synergy. As computing evolves, higher levels come to drive success. In general, the highest level of a system defines its success, e.g. social networks need a community to succeed. If no community forms, it doesn't matter how easy to use, fast or reliable the software is. Lower levels are essential to avoid failure, but higher levels are essential to success.

| Level | Requirements | Errors |
|---|---|---|
| Community | Reduce community overload, clashes, increase productivity, synergy, fairness, freedom, privacy, transparency. | Unfairness, slavery, selfishness, apathy, corruption, lack of privacy. |
| Personal | Reduce cognitive overload, clashes, increase meaning transfer efficiency. | User misunderstands, gives up, is distracted, or enters wrong data. |

| | | |
|---|---|---|
| Informa-tional | Reduce information overload, clashes, increase data processing, storage, or transfer effeciency | Processing hangs, data storage full, network overload, data conflicts. |
| Mechanical | Reduce physical heat or force overload. Increase heat or force efficiency. | Overheating, mechanical fractures or breaks, heat leakage, jams. |

TABLE 24.2: Computing errors by system level.

Conversely, *any level can cause failure*, e.g. it doesn't matter how high community morale is if the hardware fails, the software crashes or the interface is unusable. An STS fails if its hardware fails, if its program crashes or if users can't figure it out. Hardware, software, personal and community failures are all computing errors (Table 24.2). The one thing they have in common is that *the system fails to perform,* and in evolution, what doesn't perform doesn't survive.

When computing was just technology, it only failed for technical reasons, but now it is socio-technology; it can also fail for social reasons. Technology is hard, but society is soft. That the soft should direct the hard seems counter-intuitive, but trees grow at their soft tips not their hard base. As a tree trunk doesn't direct its expanding canopy, so today's social computing was undreamt of by its technical base.

## 24.1.9 Design combinations


Figure 24.7.A: Remote controls for Apple products are good examples of HCI Design.

*Courtesy of Ocrho. Copyright: pd (Public Domain (information that is common property and contains no original authorship)).*


Figure 24.7.B: Remote controls for televisions are not.

Design fields combine different requirements and design levels, as in Table 24.3:

1. *Ergonomics* is the design of safe and comfortable machines for people. To design technology to human body needs like posture and eye-strain merges biology and engineering.

2. *Object design,* as defined by Norman, applies psychological needs to mechanical design (Norman, 1990), e.g. a door's design affects whether it is pushed or pulled. An affordance is a physical object design that cues its use, as a button cues pressing. Physical systems designed to human requirements work better. In World War II, planes crashed until engineers designed cockpit controls to the cognitive needs of pilots as follows (with computing examples):
    1. Put the control by the thing controlled, e.g. a handle on a door (context menus).
    2. Let the control "cue" the required action, e.g. a joystick (a 3D screen button).
    3. Make the action/result link intuitive, e.g. press a joystick forward to go down, (press a button down to turn on).
    4. Provide continuous feedback, e.g. an altimeter, (a web site breadcrumbs line).
    5. Reduce mode channels, e.g. altimeter readings, (avoid edit and zoom mode confusions).
    6. Use alternate sensory channels, e.g. warning sounds, (error beeps).
    7. Let pilots "play", e.g. flight simulators, (a system sandbox).

3. *Human computer interaction* applies psychological requirements to software design. Usable interfaces respect cognitive principles, e.g. by the nature of human attention, users don't usually read the entire screen. HCI turns psychological needs into IT designs as architecture turns buyer needs into house designs. Compare Steve Jobs' IPod to a television re-

mote (Figure 24.7). Both do the same job[20] but one is a cool tool and the other a mass of buttons. One was designed to engineering requirements and the other to human needs. Which then performs better?

4. *Fashion* is the social need to look good applied to object design. In computing, a mobile phone can be a fashion accessory, just like a hat or handbag. Its role is to impress, not just to function. Aesthetic criteria apply when people buy mobile phones to be trendy or fashionable, so color is as important as battery life in mobile phone selection.

5. *Socio-technology,* the social design of information technology, applies social requirements to software design. Anyone online can see the power of socio-technology but most see it as an aspect of their specialty. Sociologists study society as if it were apart from physicality, which it is not. Technologists study technology as it were apart from community, which it is not. Only socio-technology studies how the social links to the technical, as a new discipline.

| Field | Target | Requirements | Example |
|---|---|---|---|
| STS | IT | Community ... | Wikipedia, YouTube, E-bay |
| Fashion | Accessory | Community ... | Mobile phone as an accessory |
| HCI | IT | Personal ... | Framing, border contrast, richness |
| Design | Technology | Personal ... | Keyboard, mouse |
| Ergonomics | Technology | Biological ... | Adjustable height screen |

TABLE 24.3: Design fields by target and requirement levels.

20. In fact the IPod does more

In Figure 24.8, higher level requirements filter down to affect lower level operation and design. This *higher affects lower* principle is that higher levels directing lower ones improves system performance. Any level requirement can translate down, e.g. communities require agreement to act, which at the citizen level gives norms, at the informational level laws and at the physical level cultural events. The same applies online, e.g. online communities make demands of netizens[21] as well as hardware. STS design then is about having it all: reliable devices, efficient code, intuitive interfaces and sustainable communities.

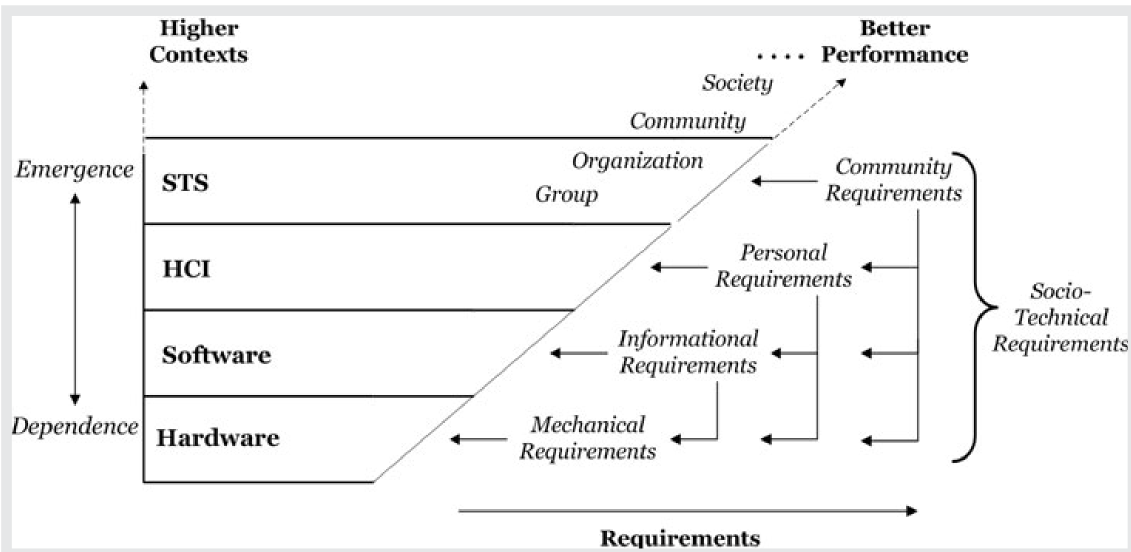



Figure 24.8: Computing requirements cumulate.



In physical society, over thousands of years, families formed tribes, tribes formed city states, city-states formed nations states, and nations formed nations of nations, each with more complex social structures (Diamond, 1998). The social level in Figure 24.8 isn't just one step, as social units can form bigger social units[22] to get new requirements (Whitworth and Whitworth, 2010).

21. For example, online 'netiquette', see: http://www.kent.edu/dl/technology/etiquette.cfm

22. A social unit of analysis can be a person, a dyadic friendship, a group, a tribe, etc.

### 24.1.10 The flower of computing

The evolution of computing involves four main specialties (Figure 24.9), but pure engineers see only mechanics, pure computer scientists only information, pure psychologists only cognitions and pure sociologists only social structures. So computing as a whole isn't pure, yet this hybrid is the future because performance isn't about purity, as practitioners understand (Raymond, 1999).

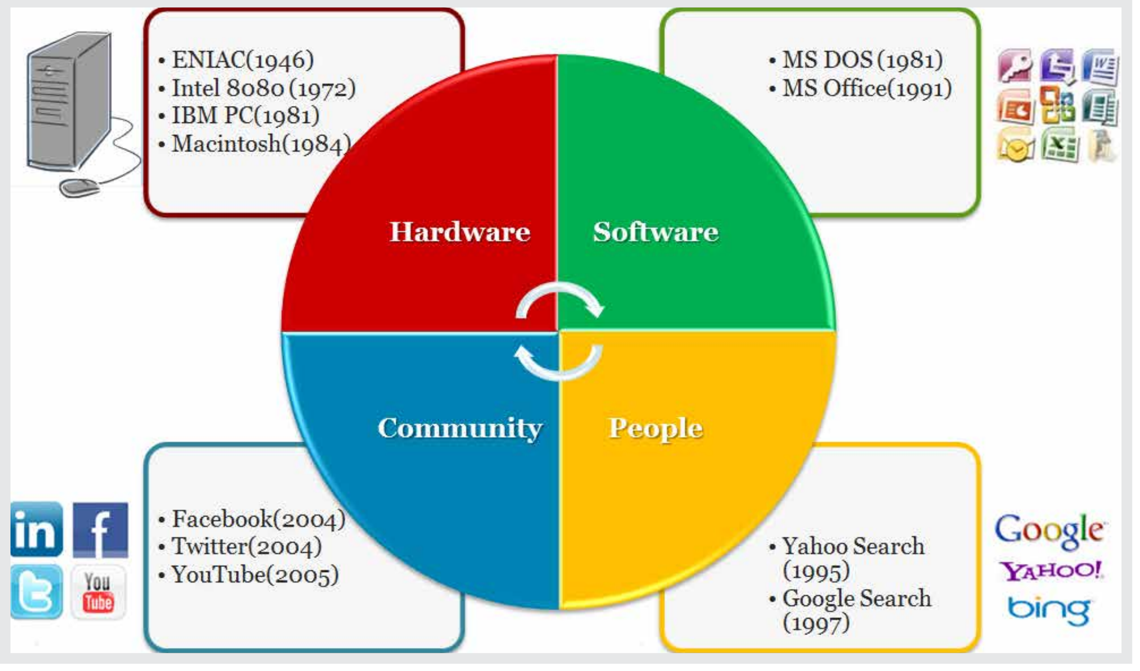



Figure 24.9: The four stages of computing.

*Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.*

The kingdom of computing is a realm divided, as academics specialize to get publications, grants and promotions (Whitworth and Friedman, 2009). Specialties guard their knowledge in journal castles with jargon walls, like medieval fiefdoms, but in doing so hold hostage knowledge, that by its nature should be free. This division also disguises and limits the growth of computing. Every day more people use more computers to do more things in more ways but computing staff rarely get critical mass, because engineering, computer science, health[23], business, psychology, mathematics and education all compete for the computing crown[24]. A realm divided is weak, and will get weaker if music, art, journalism, architecture etc. also set up outposts. Computing faculty scatter over the academic landscape like the tribes of Israel, some in engineering, some in computer science, some in health, etc. *Yet we are one*. Mathematics split up like this would be equally dilute.

The flower of computing is borne of many disciplines but belongs to none. It is a new discipline in itself (Figure 24.10). For it to bear research fruit, its academic parents must set it free. Let us trade knowledge not dominate it. Using different terms, models and theories for the same subject invites confusion. Universities that split computing research into small groups, isolated by discipline boundaries, distance themselves from its multi-disciplinary future. Until computing research becomes one, computing theory will remain as it is now - decades behind computing practice.

23. Health even created its own computing field of informatics, with separate journals, conferences and courses, to meet its non-engineering and non-business computing needs

24. Computing is the Afghanistan of academia, often invaded but never conquered. It should be the Singapore, a knowledge trade centre

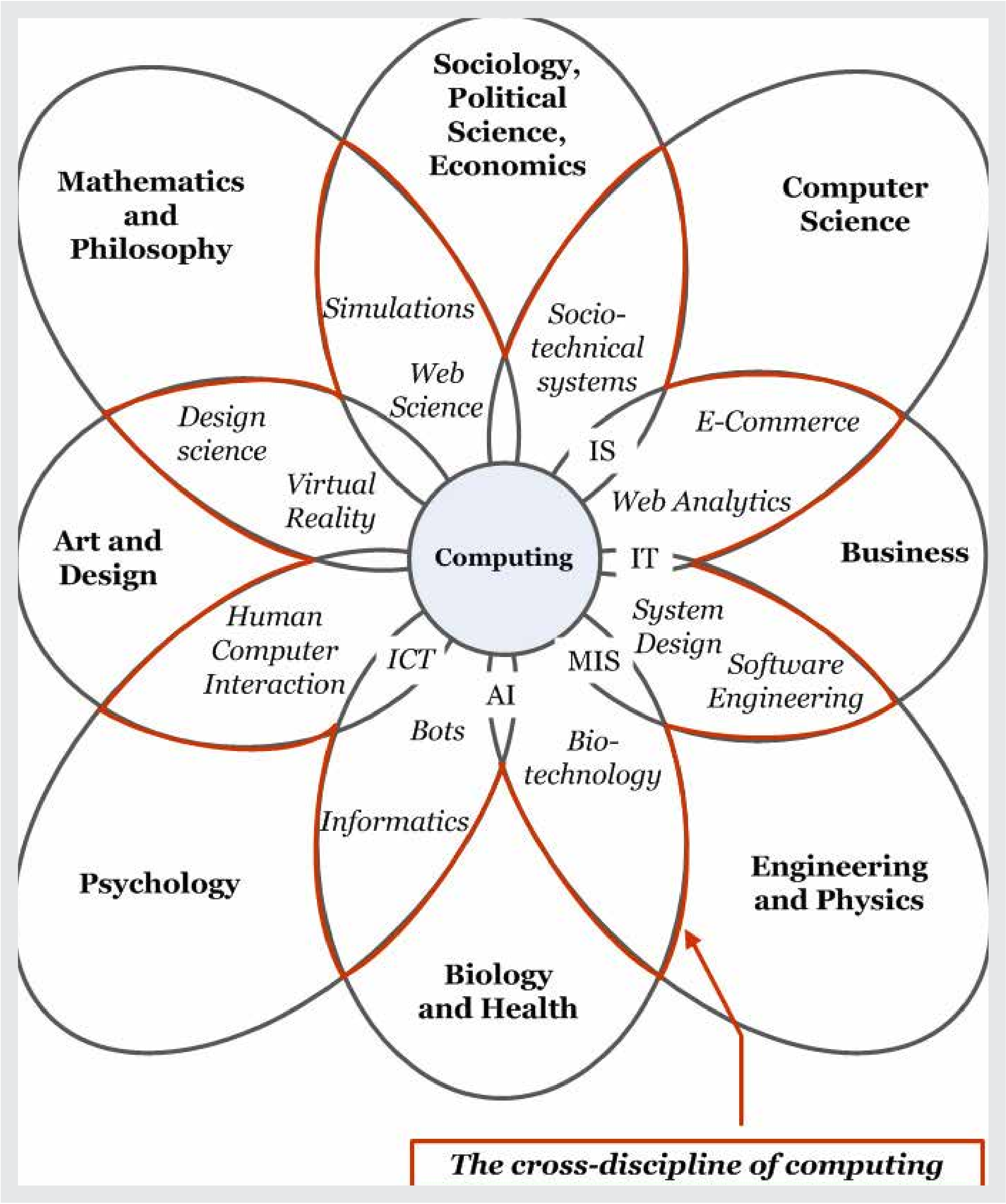



Figure 24.10: The flower of computing.

### 24.1.11 Discussion questions

Research selected questions from the list below. If you are reading this chapter as part of a class - either at university or a commercial course - you can research these questions in pairs and report back to the class, with reasons and examples.

1. How has computing evolved since it began? Is it just faster machines and better software? What is the role of hardware companies like IBM and Intel in modern computing?

2. How has the computing business model changed as it evolved? Why does selling software make more money than selling hardware? Can selling knowledge make even more money? What about selling friendships? Can one sell communities?

3. Is a kitchen table a technology? Is a law a technology? Is an equation a technology? Is a computer program a technology? Is an information technology (IT) system a technology? Is a person an information technology? Is an HCI system (person plus computer) an information technology? What, exactly, *isn't* a technology?

4. Is any set of people a community? How do people form a community? Is a socio-technical system (an online community) any set of HCI systems? How do HCI systems form an online community?

5. Is computer science part of engineering or of mathematics? Is human computer interaction (HCI) part of engineering, computer science or psychology? Is socio-technology part of engineering, computer science, psychology or one of the social sciences?[25]

6. In an aircraft, is the pilot a person, a processor, or a physical object? Can one consistently divide the aircraft into human, computer and mechanical parts? How can one see it?

25. Like, sociology, history, political science, anthropology, ancient history, etc.

7. What is the reductionist dream? How did it work out in physics? Does it recognize computer science? How did it challenge psychology? Has it worked out in any discipline?

8. How much information does a physical book, that is fixed in *one* way, by definition, have? If we say a book "contains" information, what is assumed? How is a book's information generated? Can *the same* physical book "contain" different information for different people? Give an example.

9. If information is physical, how can data compression put the same information in a physically smaller signal? If information is not physical, how does data compression work? Can one encode more than one semantic stream into one physical message? Give an example.

10. Is a bit physical "thing"? Can you see or touch a bit? If a signal wire sends a physical "on" value, is that always a bit? If a bit isn't physical, can it exist without physicality? How can a bit *require* physicality but not itself *be* physical? What creates information, if it is not the mechanical signal?

11. Is information concrete? If we can't see information physically, is the study of information a science? Explain. Are cognitions concrete? If we can't see cognitions physically, is the study of cognitions (psychology) a science? Explain. What separates science from imagination if it isn't physicality?

12. Give three examples of other animal species who sense the world differently from us. If we saw the world as they do, would it change what we *do?* Explain how *seeing* a system differently can change how it is *designed*. Give examples from computing.

13. If a $1 CD with a $1,000 software application on it is insured, what do you get if it is destroyed? Can you insure something that is not physical? Give current examples.

14. Is a “mouse error” a hardware, software or HCI problem? Can a mouse’s hardware affect its software performance? Can it affect its HCI performance? Can mouse software affect HCI performance? Give examples in each case. If a wireless mouse costs more and is less reliable, how is it better?

15. Give three examples of a human requirement giving an IT design heuristic. This is HCI. Give three examples of a community requirement giving an IT design heuristic. This is STS.

16. Explain the difference between a hardware error, a software error, a user error and a community error, with examples. What is the common factor here?

17. What is an application sandbox? What human requirement does it satisfy? Show an online example.

18. Distinguish between a personal requirement and community requirement in computing. Relate to how STS and HCI differ and how socio-technology and sociology differ. Why can’t sociologists or HCI experts design socio-technical systems?

19. What in general to people do if their needs aren’t met by a physical situation? What do users do if their needs aren’t met online? What is the difference? What do citizens of a physical community do if it doesn’t meet their needs? What about an online community? Again, what is the difference? Give specific examples to illustrate.

20. According to Norman, what is ergonomics? What is the difference between ergonomics and HCI? What is the difference between HCI and STS?

21. Give examples of: Hardware meeting engineering requirements. Hardware meeting Computer Science requirements. Software meeting CS

requirements. Hardware meeting psychology requirements. Software meeting psychology requirements. People meeting psychology requirements. Hardware meeting community requirements. Software meeting community requirements. People meeting community requirements. Communities meeting their requirements. Which are computing design

22. Why is an IPod so different from TV or video controls? Which is better and why? Why has TV remote design changed so little in decades? If TV and the Internet compete for the hearts and minds of viewers, who will win?

23. How does an online friend differ from a physical friend? Can friendships transcend physical and electronic interaction architectures? Give examples. How is this possible?

24. Why do universities spread computing researchers across many disciplines? What is a cross-discipline? What past cross-disciplines became disciplines. Why is computing a cross-discipline?

## 24.2 PART 2: DESIGN SPACES

""All my cuts are the best" (said by a butcher to a housewife who asked him for the best cuts)"

The previous section reviewed computing system levels, this one reviews constituent parts.

### 24.2.1 The elephant in the room

The beast of computing has regularly defied pundit predictions. Key advances like the cell-phone (Smith et al, 2002) and open-source development (Campbell-Kelly, 2008) weren't predicted by the experts of the day, though the signs were there for all to see. As experts pushed media-rich systems, lean text chat, blogs, texting and wikis took off. Even today, people with rich video-phones still text. Google's simple white screen scooped the search engine field not Yahoo's multi-media graphics. In gaming, the innovation was social gaming not virtual reality helmets. Investors in Internet bandwidth lost money when the future wasn't all video.

In computing, that *practice leads but theory bleeds* has a long history. Over thirty years ago, paper was declared "dead", by the electronic paperless office (Toffler, 1980). Yet today, paper is used more than ever before. James Martin saw program generators replacing programmers, but today, we still have a programmer shortage. A "leisure society" was supposed to arise as machines took over our work, but today, we are less leisured than we ever were (Golden and Figart, 2000). The list goes on: email was supposed to be for routine tasks, the Internet was supposed to collapse without central control, video was supposed to replace text, teleconferencing was supposed to replace air travel, AI smart-help was supposed to replace help-desks, and so on.

We get it wrong time and again, because *computing is the elephant in our living room*. We can't see it because it is too big. In the story of the blind men and the elephant, one grabbed its tail and found it like a rope and bendy, another took a leg and declared it fixed like a pillar, a third felt an ear and thought it like a rug and floppy, while the last seized the trunk, and found it like a pipe but very strong (Sanai, 1968). Each saw a part but none saw the whole. How can one see an elephant by analyzing its toenails?[26]

### 24.2.2 Design requirements

To design a system is to find problems early, e.g. a misplaced wall on an architect's plan can be moved by the stroke of a pen, but design needs performance requirements, like efficiency. *Requirements engineering* analyzes stakeholder needs, to

26. Yet one can see its genome in any cell, because in nature, each cell has the information to regenerate the elephant

specify what a system must do for them to sign off on the end product. It is basic to system design:

> "*The primary measure of success of a software system is the degree to which it meets the purpose for which it was intended. Broadly speaking, software systems requirements engineering (RE) is the process of discovering that purpose…*"
>
> -- *Nuseibeh and Easterbrook, 2000: p. 1*

A requirement can be a particular value (e.g. uses SSL), a range of values (e.g. less than $100), or a criterion scale (e.g. is secure). Given a system's requirements, designers can build it, but for computing, the literature can't agree on what they are. One text has usability, repairability, security and reliability (Sommerville, 2004, p. 24) but the ISO 9126-1 quality model has functionality, usability, reliability, efficiency, maintainability and portability (Losavio et al, 2004). Berners-Lee made scalability a World Wide Web criterion (Berners-Lee, 2000) while others stress open standards between systems (Gargaro et al, 1993). Business criteria are cost, quality, reliability, responsiveness and conformance to standards (Alter, 1999), but software architects prefer portability, modifiability and extendibility (de Simone and Kazman, 1995). Others espouse flexibility (Knoll and Jarvenpaa, 1994) and privacy (Regan, 1995). On the issue of what computer systems need to succeed, the literature is at best confused. This gives what developers call the *requirements mess* (Lindquist, 2005), that has ruined many a software project. It is the problem that agile methods address.

In current theories, each specialty sees only itself. Security specialists see *security* as availability, confidentiality and integrity (OECD, 1996), so to them, reliability is part of security. Reliability specialists see *dependability* as reliability, safety, security and availability (Laprie and Costes, 1982), so to them security is part of a general reliability concept. Yet both can't be generally true. Similarly, a *usability* review finds functionality and error tolerance part of usability (Gediga et al, 1999) while a *flexibility* review finds scalability, robustness and connectivity aspects of flexibility (Knoll and Jarvenpaa, 1994). In academia, each specialty expands to fill the theory space around it.

Yet there is recognition that no specialty is the be all or end all:

> "*The face of security is changing. In the past, systems were often grouped into two broad categories: those that placed security above all other requirements, and those for which security was not a significant concern. But … pressures … have forced even the builders of the most security-critical systems to consider security as only one of the many goals that they must achieve.*"
>
> -- *Kienzle and Wulf, 1998*: p5

Analyzing performance goals in isolation is giving diminishing returns.

### 24.2.3 Design spaces

Architect Christopher Alexander observed that vacuum cleaners with powerful engines and more suction were also heavier, noisier and cost more (Alexander, 1964). One performance criterion has a best point, but two criteria, like power and cost, give a best line. The *efficient frontier* of two performance criteria is the maximum of one for a value of the other (Keeney and Raiffa, 1976). A system design is choosing a many value point in a multi-dimensional *design space,* of many combinations. So there are many "best" points, e.g. a cheap, heavy but powerful vacuum cleaner, or light, expensive and powerful one (Figure 24.11). The efficient frontier of a design space is a *surface* of "best" combinations[27]. Advanced system performance is not a one dimensional ladder to excellence, but a station with many trains to many destinations.

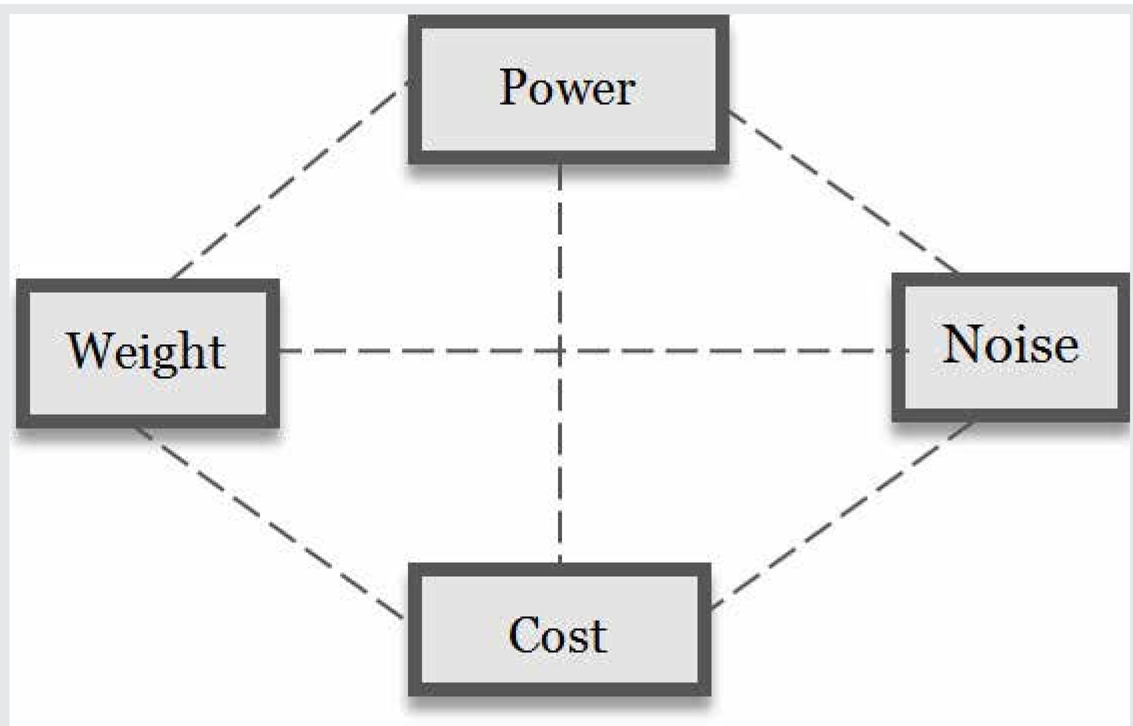



FIGURE 24.11: A vacuum cleaner design space.



Designing in a multi-dimensional space gives many "best" points, so nature has no best animal. Successful life includes flexible viruses, reliable plants, social insects and powerful tigers, with the latter endangered. In evolution, not just

27. Not all possible criterion combinations may be achievable, e.g. a light, cheap and powerful vacuum cleaner

the strong are fit and over specialization can even lead to extinction. Likewise, computing has no "best". If computer performance was just about processing we would all want supercomputers, but laptops with less power perform better for some (David et al, 2003). Blindly adding software functions gives *bloatware*[28], applications full of features that no-one uses.

Design is then the art of reconciling many requirements in a system form, e.g. a quiet, reliable, cheap and powerful vacuum cleaner. It is the *innovative synthesis of a performance form in a requirements space* (Alexander, 1964). It isn't one dimensional, e.g. Berners-Lee chose HTML for the World Wide Web for its flexibility (across platforms), reliability *and* usability (easy to learn). An academic conference rejected his WWW proposal because HTML was inferior to SGML (Standard Generalized Markup Language). Specialists saw their specialty, not system performance. Even after the World Wide Web's phenomenal success, their blindness remained:

> "*Despite the Web's rise, the SGML community was still criticising HTML as an inferior subset … of SGML*"
>
> -- *Berners-Lee, 2000: p96*

What has changed since academia found the World Wide Web "inferior"? Not a lot. If it is any consolation, an equally myopic Microsoft also found it "unprofitable". In system design, a focus on any one criterion gives diminishing returns, whether it is functionality, security (OECD, 1996), extendibility (Simone and Kazman, 1995), privacy (Regan, 1995), usability (Gediga et al., 1999) or flexibility

28. Also called featuritis or scope creep

(Knoll and Jarvenpaa, 1994). Improving one aspect alone can even *reduce* performance, i.e. "bite back" (Tenner, 1997), e.g. a network so secure that no-one uses it. Advanced system performance does not result from one dimensional design.

### 24.2.4 Non-functional requirements

In traditional requirements engineering, criteria like usability are quality requirements that affect functional goals but can't stand alone (Chung et al, 1999). For decades, these non-functional requirements (NFRs), or "-ilities", were considered second class requirements. They defied categorization, except to be non-functional. How exactly they differed from functional goals was never made clear (Rosa et al, 2001), yet most modern systems have more lines of interface, error and network code than functional code, and increasingly fail for "unexpected" non-functional reasons[29] (Cysneiros and Leite, 2002, p. 699).

The logic is that NFRs like reliability can't exist without functionality, so are subordinate to it. Yet by the same logic, functionality can't exist without reliability, e.g. a car that won't start has no speed function, nor does a car that is stolen or can't be driven. NFRs don't just modify performance they define it. In nature, functionality isn't the only key to success, e.g. viruses hijack the functionality of other system's. Functionality differs from other system requirements only in being more obvious *to us*. It is really just one of many requirements. The distinction between functional and non-functional requirements is our bias, like seeing the sun going round the earth because we are on the earth.

### 24.2.5 Constituent parts

In general systems theory, any system consists of:

1. Parts, and
2. Interactions.

29. Hardly surprising if we define NFRs to be less important.

But are software parts lines of code, variables or sub-programs? Let a system's *elemental parts* be those not formed of other parts. A mechanic stripping a car stops at the bolt element, as to decompose it further gives atoms, which are no longer mechanical. Each level has a different elemental part: physics has quantum strings, information has bits, psychology has qualia, and society has citizens (Table 24.4). Elemental parts then form complex parts as bits form bytes.

| Level | Elemental part | Other parts |
|---|---|---|
| *Community* | Citizen | Friendships, groups, organizations, societies. |
| *Personal* | Qualia | Cognitions, attitudes, beliefs, feelings, theories. |
| *Informa-tional* | Bit | Bytes, records, files, commands, databases. |
| *Physical* | Quantum strings? | Quarks, electrons, nucleons, atoms, molecules. |

Table 24.4: System parts by level.

Let a system's *constituent parts* be those that interact to form the system but are *not* part of other parts (Esfeld, 1998). So, disconnecting a car entirely gives elemental parts not constituent parts, e.g. a bolt on a wheel isn't a constituent because it is part of the wheel.

To say a body is composed of cells ignores its structure: how elemental parts form constituent parts. Only in system heaps, like a pile of sand, are elemental parts also constituent parts. The body's constituent parts are the digestive system, the respiratory system, etc, not its cells. Just sticking together arbitrary physical parts, like head, arms, and legs, gives the *Frankenstein effect*[30] (Tenner, 1997).

### 24.2.6 Holism and specialization

The performance of a system of parts that interact isn't defined by decomposition alone. Even simple parts, like air molecules, can interact strongly to form a chaotic system like the weather (Lorenz, 1963). Gestalt psychologists called the whole being more than its parts *holism*, as a curve is just a curve but in a face becomes a "smile". Holism is how system parts change by interacting with others. Holistic systems are individualistic, because changing one part, by its interactions, can cascade to change the whole system drastically. People rarely look the same because one gene change can change everything. The brain is also holistic - one thought can change everything you know.

Yet a system's parts needn't be simple. The body began as one cell, a zygote, that divided into all the cells of the body, including liver, skin, bone and brain cells[31]. Likewise in early societies most people did most things, but today we have millions of specialist jobs. A system's *specialization*[32] is the degree its parts differ in form and action, especially constituent parts.

Holism (complex interactions) and specialization (complex parts) are hallmarks of evolved systems, giving both levels and constituent specializations.

### 24.2.7 General performance requirements

Requirements engineering aims to define a system's purposes. If levels and constituent specializations change those purposes, how can requirements engineer-

30. Dr Frankenstein made a human being by putting together the best of each individual body part he could find in the graveyard. The result was a monster.

31. Deciphering the human genome gave the pieces of the genetic puzzle, not how they connect

32. Specialization is also called differentiation

ing succeed? The answer proposed here is to take the view *of the system itself,* specifying requirements for different levels and constituent specializations. How these are reconciled is then the art of system design.

A system interacts with its environment to *perform,* i.e. to gain value and avoid loss in order to survive. In Darwinian terms, what doesn't survive fails and what does succeeds. So a system needs a *boundary* to exist apart from the world and an internal *structure* to support and manage that existence. It needs *effectors* to act upon the environment around it and *receptors* to monitor the world for risks and opportunities.

| Constituent | Requirement | Definition |
|---|---|---|
| *Boundary* | Security | *To deny unauthorized entry, misuse or takeover by other entities.* |
| | Extendibility | *To attach to or use outside elements as system extensions.* |
| *Structure* | Flexibility | *To adapt system operation to new environments* |
| | Reliability | *To continue operating despite system part failure* |
| *Effector* | Functionality | *To produce a desired change on the environment* |
| | Usability | *To minimize the resource costs of action* |
| *Receptor* | Connectivity | *To open and use communication channels* |
| | Privacy | *To limit the release of self information by any channel* |

Table 24.5: System performance requirements by constituent specialty.

So as cells evolved they first got a boundary membrane, then organelle and nuclear structures for support and control, then eukaryotic cells evolved flagella to move and protozoa got photo-receptors (Alberts et al, 1994). We also have a skin boundary, metabolic and brain structures, muscle effectors and sense receptors, like the eye. Computers also have a case boundary, a motherboard internal structure, printer or screen effectors and keyboard or mouse receptors. Four constituent specializations by risk and opportunity goal options gives eight *performance requirements* (Table 24.5). The details are as follows:

1. *Boundary* constituents manage the system boundary. They can be designed to deny outside things entry (security) or to use them (extendibility). In computing, virus protection is security and system add-ons are extendibility (Figure 24.12). In people, the immune system gives biological security and tool-use illustrates extendibility.

2. *Structure* constituents manage internal operations. They can be designed to limit internal change to reduce faults (reliability), or to allow internal change to adapt to outside changes (flexibility). In computing, reliability reduces and recovers from error and flexibility is the system preferences that allow customization. In people, reliability is the body fixing a cell «error» that might cause cancer, while the brain learning illustrates flexibility.

3. *Effector* constituents manage environment actions, so can be designed to maximize effects (functionality) or minimize resource use (usability). In computing, functionality is the menu functions, while usability is how easy they are to use. In people, functionality gives muscle effectiveness and usability is metabolic efficiency.

4. *Receptor* constituents manage signals to and from the environment, so can be designed to open communication channels (connectivity) or close

them (privacy). Connected computing can download updates or chat online, while privacy is the power to disconnect or log off. In people, connectivity is conversing and privacy is the legal right to be left alone. In nature, privacy is camouflage, and the military calls it stealth.

Every system is somehow created, which takes effort both for applications that are built or organisms that are born. A system's ability to reproduce is important but outside the current scope, as apart from virus programs, few computer systems do this.

These general system criteria map well to current terms (Table 24.6). They apply at any level, but as what is exchanged changes, so do the names used:

1. *Hardware systems exchange energy*. So "functionality" is power, i.e. hardware with high CPU cycle or disk read-write rates. "Usable" hardware uses less power for the same result, e.g. mobile phones that last longer. Reliable hardware is rugged enough to work if you drop it, and flexible hardware is mobile to still work if you move around, i.e. change environments. Secure hardware blocks physical theft, e.g. by laptop cable locks, and extendible hardware has ports for peripherals to be attached. Connected hardware has wired or wireless links and private hardware is tempest proof i.e. it doesn't physically leak energy.

2. *Software systems exchange information*. Functional software has many ways to process information, while "usable" software uses less CPU processing ("lite" apps). Reliable software avoids errors or recovers from them quickly. Flexible software is operating system platform independent. Secure software can't be corrupted or overwritten. Extendible software can access OS program library calls. Connected software has protocol "handshakes" to open read/write channels. Private software can encrypt information so others can't see it.

3. *HCI systems exchange meaning*, including ideas, feelings and intents. In functional HCI the human computer pair is effectual, i.e. meets the task goal. Usable HCI requires less intellectual, affective or conative[33] effort, i.e. is intuitive. Reliable HCI avoids or recovers from unintended user errors by checks or undo choices — the web Back button is an HCI invention. Flexible HCI lets users change language, font size or privacy preferences, as each person is a new environment to the software. Secure HCI avoids identity theft by user password. Extendible HCI lets users use what others create, e.g. mash-ups and third party add-ons. Connected HCI communicates with others, while privacy includes not getting spammed or being located on a mobile device.

Each level applies the same ideas to a different system view. The community level is covered later.

| GSR Criterion | Related Criteria |
|---|---|
| *Functionality* | Effectualness, capability, usefulness, effectiveness, power, utility. |
| *Usability* | Ease of use, simplicity, user friendliness, efficiency, accessibility. |
| *Extendibility* | Openness, interoperability, permeability, compatibility, standards. |
| *Security* | Defense, protection, safety, threat resistance, integrity, inviolable. |
| *Flexibility* | Adaptability, portability, customizability, plasticity, agility, modifiability. |

33. Conative refers to the will; affective refers to the emotions; while intellectual refers to thoughts. All are cognitions that form from perceptions.

| | |
|---|---|
| *Reliability* | Stability, dependability, robustness, ruggedness, durability, availability. |
| *Connectivity* | Networkability, communicability, interactivity, sociability. |
| *Privacy* | Tempest proof, confidentiality, secrecy, camouflage, stealth, encryption. |

Table 24.6: Related performance criteria.


Figure 24.12: Mozilla/Firefox add-ons.

*Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.*

### 24.2.8 A general system design space

The above gives the general system design space of Figure 24.13, where for a particular system:

- *The area* is the overall performance requirements met, i.e. performance in general.
- *The shape* is the requirement weights, defined by the environment.
- *The lines* are design requirement "tensions" (see below).

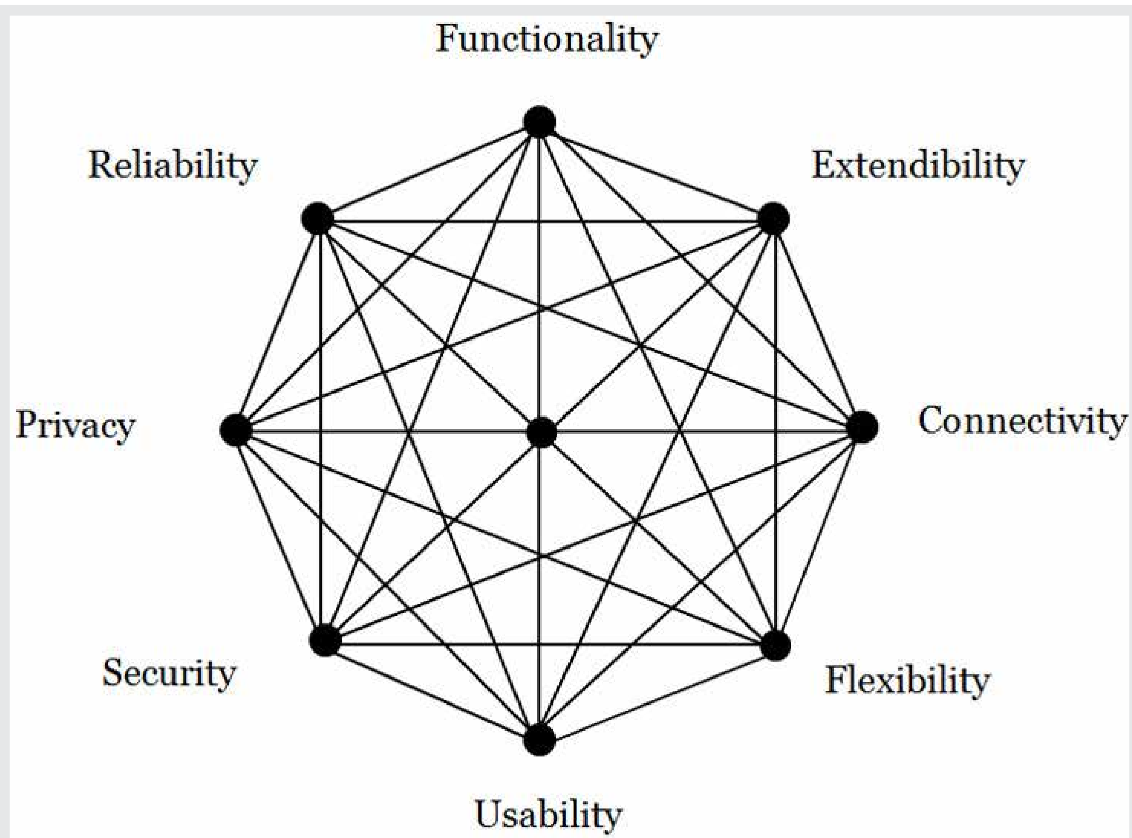



FIGURE 24.13: A general system performance design space.

This space has active requirement that enhance opportunities[34] and passive ones that reduce risks[35], where taking opportunities is as important in performance as reducing risk (Pinto, 2002). Criteria weights vary by environment, so security is more important in threat environments and extendibility is better in opportunity environments (Whitworth et al, 2008). These performance criteria are general because they have no inherent contradictions, e.g. a bullet-proof plexi-glass room can be secure but not private, while encrypted files can be private but not secure. Reliability provides services but security denies them (Jonsson, 1998), so a system can be reliable but insecure, unreliable but secure, unreliable and insecure or reliable and secure. Functionality needn't deny usability (Borenstein and Thyberg, 1991) or connectivity privacy. C*ross-cutting requirements* (Moreira et al, 2002) can be reconciled by innovative design if they are logically modular, so one can get both.

### 24.2.9 Design tensions and innovations

A design tension is when satisfying one design requirement denies another. Applying different requirements to the same constituent gives a design tension, e.g. castle walls that protect against attacks but need a gate to receive supplies, or computers impenetrable to virus attacks that need plug-in software hooks. These contrasts are not anomalies, but built into the nature of systems.

Design tensions begin slack for new systems, but increase as performance improves. Eventually, like stretched rubber bands, the system becomes so "tight" that advancing one requirement can easily pull back another, or more than one. In the *version 2 paradox*, development effort spent improving a successful product can decrease its performance!

To expand a performance web, one can't just pull one corner, e.g. in 1992, Apple CEO Sculley introduced the hand held Newton, claiming that portable

34. Functionality, flexibility, extendibility and connectivity

35. Security, reliability, privacy and usability.

computing was the future. We now know he was right, yet in 1998 Apple dropped the line due to poor sales. The Newton's small screen made data entry hard, i.e. the portability gain was nullified by a usability loss. Only when Palm's Graffiti language improved handwriting recognition, did the personal digital assistant market revive. Sculley's innovation was only half the answer - the other half was resolving the usability problems created by increasing flexibility. Innovative design must meet specialist requirements and resolve design tensions.

### 24.2.10 Project development

| Constituent | Code | Requirement | Analysis | Testing |
|---|---|---|---|---|
| **Actions** | Application | *Functionality* | Task | Business |
| | Interface | *Usability* | Usability | User |
| **Interactions** | Authorization | *Security* | Threat | Penetration |
| | Plug-ins | *Extendibility* | Standards | Compatibility |
| **Changes** | Error recovery | *Reliability* | Stress | Load |
| | Preferences | *Flexibility* | Contingency | Situation (Beta) |
| **Interchanges** | Network | *Connectivity* | Channel | Communication |
| | Rights | *Privacy* | Legitimacy [36] | Community |

TABLE 24.7: Project development specializations by constituent.

36. See the next section. Legitimacy analysis specifies social interaction requirements.

The days when programmers could list a system's functions then just code them are gone, if they ever existed. Today, design involves not only many specialties but also their interaction. A system development could involve up to eight specialists, with distinct requirements, analysis and testing (Table 24.7). Smaller systems might have four (actions, interactions, changes and interchanges), two (opportunities and risks) or just one (performance). Design tensions are reduced by agile methods, where specialists talk to each other and stakeholders, but system development also needs innovators, people to cut across specialist boundaries to resolve cross-cutting design tensions.

### 24.2.11 Discussion questions

Research selected questions from the list below. If you are reading this chapter as part of a class - either at university or a commercial course - you can research these questions in pairs and report back to the class, with reasons and examples.

1. What three widespread computing expectations didn't happen? Why not? What three unexpected computing outcomes did happen? Why?

2. What is a system requirement? How does it relate to system design? How do system requirements relate to performance? Or to system evaluation criteria? How can one specify or measure system performance if there are many factors?

3. What is the basic idea of general systems theory? Why is it useful? Can a cell, your body, and the earth all be considered systems? Describe Lovelock's Gaia Hypothesis. How does it link to both General Systems Theory and the recent film Avatar? Is every system contained within another system (environment)?

4. Does nature have a best species? If nature has no better or worse, how can species evolve to be better? Or if it has a better and worse, why is

current life so varied instead of just the "best"?[37] Does computing have a best system? If it has no better or worse, how can it evolve? If it has a better and worse, why is current computing so varied? Which animal actually is "the best"?

5. Why did the electronic office *increase* paper use? Give two good reasons to print an email in an organization. How often do you print an email? When will the use of paper stop increasing?

6. Why wasn't social gaming predicted? Why are MMORPG human opponents better than computer ones? What condition must an online game satisfy for a community to "mod" it (add scenarios)?

7. In what way is computing an "elephant"? Why can't it be put into an academic "pigeon hole"?[38] How can science handle cross-discipline topics?

8. What is the first step of system design? What are those who define what a system should do called? Why can't designers satisfy every need? Give examples from house design.

9. Is reliability an aspect of security or is security an aspect of reliability? Can both these things be true? What are reliability and security both aspects of? What decides which is more important?

10. What is a design space? What is the efficient frontier of a design space? What is a design innovation? Give examples (not a vacuum cleaner).

11. Why did the SGML academic community find Tim Berners-Lee's WWW proposal of low quality? Why didn't they see the performance potential? Why did Microsoft also find it "of no business value"? How did the WWW eventually become a success? Given that business and academia now use it extensively,

37. Success in evolutionary terms is what survives. Over 99% of all species that ever existed are now extinct, so every species existing today is a great success. Bacteria and viruses are as evolved as us in evolutionary terms.

38. A pigeon-hole is a small space used to roost pigeons.

why did they reject it initially? What have they learned from this lesson?

12. Are NFRs like security different from functional requirements? By what logic are they less important? By what logic are they equally critical to performance?
13. In general systems theory (GST), every system has what two aspects? Why doesn't decomposing a system into simple parts fully explain it? What is left out? Define holism. Why are highly holistic systems also individualistic? What is the Frankenstein effect? Show a "Frankenstein" web site. What is the opposite effect? Why cant "good" system components just be stuck together?
14. What are the elemental parts of a system? What are its constituent parts? Can elemental parts be constituent parts? What connects elemental and constituent parts? Give examples.
15. Why are constituent part specializations important in advanced systems? Why do we specialize as left-handers or right-handers? What about the ambidextrous?
16. If a car is a system, what are its boundary, structure, effector and receptor constituents? Explain its general system requirements, with examples. When might a vehicle's "privacy" be a critical success factor? What about its connectivity?
17. Give the general system requirements for browser application. How did its designers meet them? Give three examples of browser requirement tensions. How are they met?
18. How do mobile phones meet the general system requirements, first as hardware and then as software?
19. Give examples of usability requirements for hardware, software and HCI. Why does the requirement change by level? What is "usability" on a community level?

20. Are reliability and security really distinct? Can a system be reliable but insecure, unreliable but secure, unreliable and insecure, or reliable and secure? Give examples. Can a system be functional but not usable, not functional but usable, not functional or usable, or both functional and usable? Give examples.

21. Performance is taking opportunities and avoiding risks. Yet while mistakes and successes are evident, missed opportunities and mistakes avoided aren't. Explain how a business can fail by missing an opportunity, with WordPerfect vs Word as an example. Explain how a business can succeed by avoiding risks, with air travel as an example. What happens if you only maximize opportunity? What if you only reduce risks? Give examples. How does nature both take opportunities and avoid risks? How should designers manage this?

22. Describe the opportunity enhancing general system performance requirements, with an IT example of each. When would you give them priority? Describe the risk reducing performance requirements, with an IT example of each. When would you give them priority?

23. What is the Version 2 paradox? Give an example from your experience, of software that got worse on an update. You can use a game example. Why does this happen? How can designers avoid this?

24. Define extendibility for any system. Give examples for a desktop computer, a laptop computer and a mobile device. Give examples of software extendibility, for email, word processing and game applications. What is personal extendibility? Or community extendibility?

25. Why is innovation so hard for advanced systems? Why stops a system being secure and open? Or powerful and usable? Or reliable and flexible? Or connected and private? How can such diverse requirements ever be reconciled?

26. Give two good reasons to have specialists in a large computer project team. What happens if they disagree? Why are cross-disciplinary integrators also needed?

## 24.3 PART 3: SOCIO-TECHNICAL DESIGN

"Let the social define the technical"

Social ideas like freedom seem far removed from computer code but computing today *is* social. That technology designers aren't ready, have no precedent or don't recognize social needs is irrelevant. Like a baby being born, online society is pushing forward, ready or not. And like new parents, socio-technical designers are causing it, whether they want to or not. As the World Wide Web's creator observes:

> "*... technologists cannot simply leave the social and ethical questions to other people, because the technology directly affects these matters*"
>
> *-- Berners-Lee, 2000: p124*

The online reality is that how people interact in socio-technical systems depends entirely on the software.

### 24.3.1 Designing work management

The term socio-technical was first introduced by the Tavistock Institute[39] in the late 1950's to oppose Taylorism - reducing jobs to efficient elements on assembly lines in mills and factories. Community level needs gave work-place management ideas like (Porra and Hirschheim, 2007):

1. *Congruence*. A process must match its objective - democratic results need democratic means.

39. see http://www.strategosinc.com/socio-technical.htm

2. *Minimize control*. Give employees clear goals, but let them decide how to achieve them.

3. *Local control*. Let those experiencing a problem change the system, not absent managers.

4. *Flexibility*. Without "extra" skills to handle change, specialization will precede extinction.

5. *Boundary innovation*. Innovate at the boundaries, where work goes between groups.

6. *Transparency*. Give information first to those it affects, e.g. give work rates to workers.

7. *Evolution*. Work system development is an iterative process that never stops.

8. *Lead by example*. Chinese saying: "If the General takes an egg, his soldiers will loot a village."[40]

9. *Support human needs*. Work that lets people learn, choose, feel and belong gives loyal staff.

In computing it became a call for the ethical use of technology. Yet social needs apply to technology design as well as to work management. Technology that mediates social interactions must also satisfy social needs. In the industrial revolution, "dark satanic mills" enslaved people, so technology was the enemy. Yet *people* ran those factories. It was the rich oppressing the poor, as always, with machines just letting them do it better. Technology is an effect magnifier, i.e. it isn't in itself good or evil. The people of nineteenth century Britain rejected slavery[41] but embraced

40. While Steve Jobs worked for $1 per year, other CEOs take all they can get - *simply because they can*.

41. The industrial revolution brought the feudalism myth to a head, as socialism and communism fought class slavery. Last century's technology brought the myth of world domination to a head, as a world wide peace movement fought war. The myth today's information revolution challenges is perpetual profit, the fantasy of getting something for nothing that drove Enron, Worldcom and the sub-prime

car and phone technologies. In today's information revolution we "love" technology. It is on the other side of the class war, as Twitter, Facebook and YouTube support the Arab spring. Yet the core socio-technical principle is the same:

> *"To participate in a market economy, to be willing to ship goods to distant destinations and to invest in projects that will come to fruition or pay dividends only in the future, requires confidence, the confidence that ownership is secure and payment dependable. ... knowing that if the other reneges, the state will step in..."*
>
> *-- Mandelbaum, 2002: p272*

In the middle ages, democracies weren't just unthinkable, they were also unworkable. Freeing people who aren't ready just gives anarchy and a return to autocracy, as the French revolution gave the terror then the Emperor Napoleon[42]. Yet America and England somehow got democracy, and now it is unclear why our predecessors ever settled for less. Democracies out-produce autocracies as free people do more and online is no different (Beer and Burrows, 2007). Communities perform by improving social interactions, which happens when citizens do what they should - not what they can.

### 24.3.2 Social requirements

One can't design socio-technology in a social vacuum. Fortunately, while virtual society is new people have been socializing for thousands of years. We know

meltdown. We laugh at myths of perpetual youth or perpetual motion, yet today seek perpetual profit, which is equally impossible.

42. A community can be governed in various ways. Autocracy is control by one person, aristocracy is control by an elite, plutocracy is control by the rich, democracy is control by all the citizens and anarchy is no-one in charge.

that fair communities prosper but corrupt ones don't (Eigen, 2003). Social inventions like laws, fairness, freedom, credit and contracts were bought with blood and tears (Mandelbaum, 2002), so why start anew online? Why reinvent the social wheel in cyber-space (Ridley, 2010)? Why re-learn electronically what we already know physically, if the social level in both cases is the same?

When nuclear technology magnified the physical power of war, humanity had a choice: to destroy itself physically by nuclear holocaust, or not. We didn't destroy ourselves *by choice*, not by technology, which just upped the ante. As the new bottle of information technology fills with the old wine of society, the stakes are raised again. Today's information revolution vastly increases the power to gather, store and distribute information, for good or ill (Johnson, 2001). We can be hunter-gatherers of the information age or an online civilization (Meyrowitz, 1985). Yet a stone-age society with space-age technology isn't a good mix.

In general, we are "environment blind". We don't see social environments not because they are too far away but because they are too close. As a fish is the last to see water, or a bird the air, so we can't see social environments. Yet if technology is to support civilization, it must specify its requirements. Computing can't implement what it can't specify. We live in social environments every day, but struggle to specify them[43], e.g. a shop-keeper swipes a credit card with a reading device designed to *not* store data like credit card number or pin. It is designed to the social requirement that shopkeepers don't steal from customers, even if they can. Without this, credit would collapse and a social failure, or depression, can be worse than a natural disaster. In sum, credit card readers support social trust by design.

Likewise, if online systems take and sell customer data like home address and phone for advantage, users will lose trust, and either refuse to register at all, or register with fake data, like "123 MyStreet, MyTown, NJ" (Foreman and Whitworth, 2005). The key to online privacy is not storing data. To say it will never be revealed

43. In general, entities are environment blind, e.g. fish don't see the sea or birds the air.

isn't good enough, as companies can be forced by governments or bribed by cash to reveal data. One can't be forced or bribed to give data one doesn't have. The best way to guarantee online trust is to not to store unneeded information in the first place[44].

### 24.3.3 The socio-technical gap

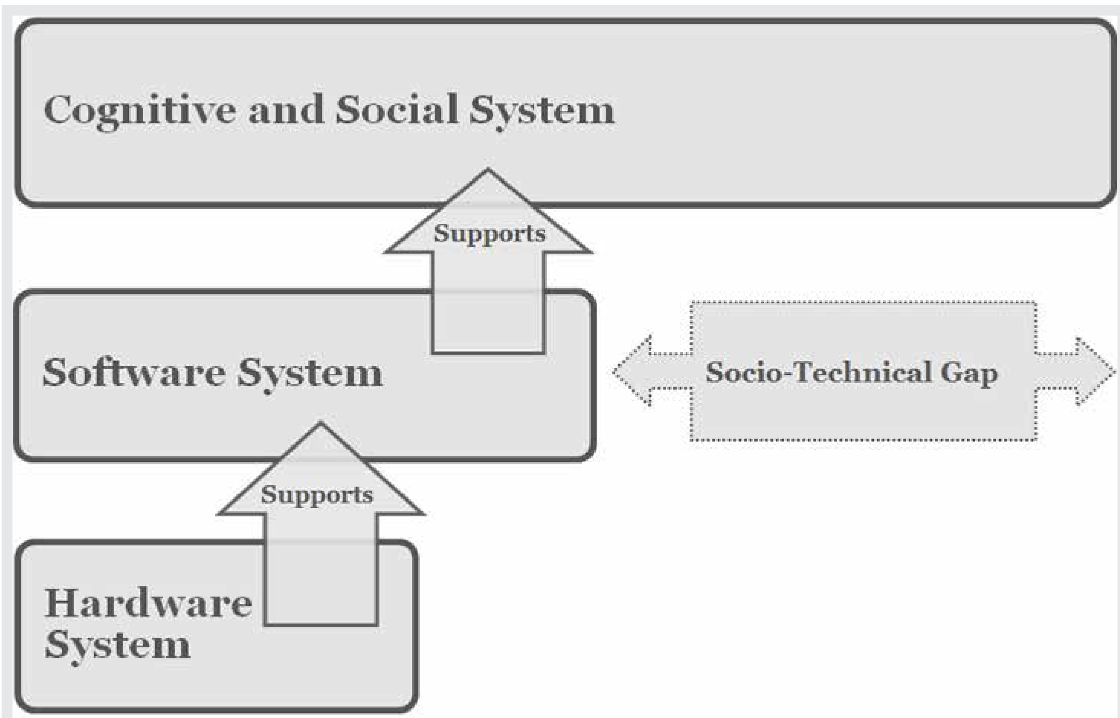



Figure 24.14: The socio-technical gap.

*Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.*

Socio-technical design is the application of community requirements to people and software and hardware. Pure technical design gives a *socio-technical gap* (Figure 24.14), between what technology supports and what people want (Ackerman, 2000), e.g. designing email to let anyone message anyone without permission gave the spam problem. Filters help on a personal level but transmitted spam

44. Trying to gather all the information you can is information greediness.

as a system problem has never stopped growing. While *inbox* spam is constant, due to filters, *transmitted* spam grew from 20% to 40% in 2002-2003 (Weiss, 2003), to 60-70% in 2004 (Boutin, 2004), to from 86.2% to 86.7% of the 342 billion emails sent in 2006 (MAAWG, 2006; MessageLabs, 2006), to 87.7% of spam in 2009 and 89.1% in 2010 (MessageLabs, 2010). A 2004 prediction that within a decade over 95% of all emails transmitted by the Internet will be spam is coming true (Whitworth and Whitworth, 2004).

Filters address spam as a *user* problem, but it is really a *community* problem. Transmitted spam uses Internet processing, bandwidth and storage whether users behind their filter walls see it or not. Only socio-technology can resolve social problems like spam, because in the “spam wars”, technology helps both sides, e.g. image spam can bypass text filters, AI can solve captchas[45], botnets can harvest web site emails, and zombie sources can send emails. So spam isn’t going away any time soon (Whitworth and Liu, 2009a).

Aliens visiting our planet might suppose our email system was build for machines, as most of the messages it transmits go from one computer (spammer) to another computer (filter), untouched by human eye.This result is not just bad luck. A communication technology isn’t a Pandora’s box, unknown until opened, *because we built it*. Spam happens when we build technologies instead of socio-technologies.

### 24.3.4 Legitimacy analysis

In politics, a *legitimate* government is seen as rightful by its citizens, i.e. accepted. In contrast, illegitimate governments need force of arms and propaganda to stay in power. By extension, *legitimate interaction* is accepted by the parties involved, who freely repeat it, e.g. fair trade. Legitimacy has been specified as: *fairness* and *public good* (Whitworth and de Moor, 2003). Physical and online citizens prefer legitimate communities because they perform better socially.

45. CAPTCHA stands for Completely Automated Public Turing test to tell Computers and Humans Apart.

In physical society, legitimacy is maintained by laws, police and prisons, that punish criminals. Legitimacy is the human concept by which judges create new laws and juries decide on never before seen cases. The higher affects lower principle applies here: communities engender human ideas like fairness, which generate informational laws, that are used to govern physical interactions. Communities affect people to create rules to direct acts that benefit the community, i.e. higher level goals drive lower level operations to improve system performance. Doing so online, applying social principles to technical systems, is socio-technical design.

Conversely, over time laws get a "life of their own" and the tail wags the dog, e.g. copyright laws designed to encourage innovators are now just a tool to perpetuate corporate profit (Lessig, 1999)[46]. Unless continuously "re-invented" at the human level, laws inevitably decay. Today's online society is a social evolution as well as a technical one. The social Internet is a move to community goals like service and freedom, so to reduce it to a hawker market place would be its devolution. So let the old ways of business, politics and academia be changed by the Internet, not the other way around.

One can't just "stretch" physical laws into cyberspace (Samuelson, 2003) because they often:

1. *Don't transfer* (Burk, 2001), e.g. what is online "trespass"?
2. *Don't apply*, e.g. what law applies to online "cookies" (Samuelson, 2003)?
3. *Change too slowly,* e.g. laws change in years but code changes in months.
4. *Depend on code* (Mitchell, 1995), e.g. anonymity means actors can't be identified.
5. *Have no jurisdiction.* U.S. law applies to U.S. soil but cyber-space isn't "in" America.

46. Disney copyrighted public domain stories like Snow White that they didn't create, then stopped others using them.

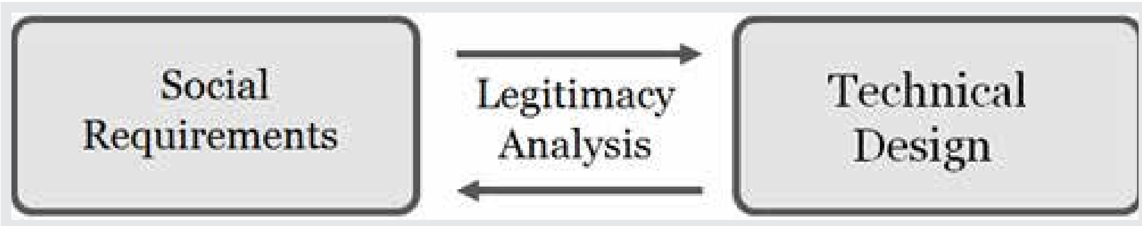



Figure 24.15: Legitimacy analysis.



There are no shortcuts here, as to repeat the past isn't progress. To get legitimacy online we must build it in again, because online code *is* law (Lessig, 1999).The software that mediates online interaction has control, e.g. any application could upload any hard drive file on your computer to any server. In itself, code could create a perfect online police state, where everyone is monitored, all "wrong" acts punished and all "undesirables" excluded, i.e. a tyranny of code.

Yet code is also an opportunity to be *better than the law,* based on legitimacy analysis (Figure 24.15). Physical justice, by its nature, operates after the fact, i.e. a person must commit a crime to be punished. With appeals, this can take years, and justice delayed is justice denied. In contrast, code as the online environment itself, acts right away. It can be designed to enable social acts not just deny anti-social ones. Socio-technical systems that are *legitimate by design* perform better socially (Whitworth and de Moor, 2003).

Saying that technology supporting social requirements, like fairness, improves system performance is the radical core of socio-technical design. So is every STS designer an application law-giver, like Moses coming down from the mountain with tablets of code instead stone? Not quite, as STS directives are to software not people. Telling people to choose rightly is the job of ethics. The job of right code, like right laws, is to allow what is legitimate, not to enforce choices on people. Socio-technical design is socializing technology not technologizing society, the higher directing the lower not the reverse.

To achieve online what laws do offline, STS developers must re-invoke legitimacy for each application. It seems hard but every citizen on jury service already interprets the "spirit of the law" for complex physical cases. STS design is the same but for application cases. That the result isn't perfect doesn't matter. Cultures have different laws and ethics but all have some laws and ethics, because some social requirements are always better than none.

Yet to build a society as one does a house is wrong. Social engineering by coercion, propaganda or indoctrination is a few enforcing their will on the many. Yet a community by definition is many people working together, so an elite few enslaving the rest isn't a community. To socially engineer a community is to treat people like bricks in a wall. It denies freedom and accountability, which are social requirements. Communities can't be "built" because their parts are actors. They just emerge as people interact.

### 24.3.5 The web of social performance requirements

Communities interact with others, using spies as "eyes", diplomats to communicate, engineers to effect, soldiers to defend, intellectuals to adapt and traders to extend, but a community can also interact *with itself,* to communicate or synergize, as follows:

1. *Productivity*. Previously, functionality was what a system can do. What communities do is to produce bridges, dams, art, science, etc. This productivity is based on citizen competence, which education systems increase. Help and FAQ systems do the same online.

2. *Synergy*. Previously, usability was less effort per result. Communities do this by synergy, by citizens giving to others[47]. Public goods like roads and hospitals are specialists giving what they do well to all. If everyone in a

47. In synergy, everyone gives so everyone gets, while if everyone takes everyone is taken from.

community specializes, and offers their services to others, all get more for less. Wikipedia is synergy, as many give a little knowledge to all get a lot.

Productivity and synergy are in tension, as one invokes competition and the other cooperation[48]. One improves what citizen "parts" do, the other how they interact. *Service* by free-good citizens reconcile them, as free citizens raise productivity and good citizens increase synergy. Free-goodness combines the invisible hand of the market and the visible hand of public good (Whitworth and Whitworth, 2010).

1. *Freedom*. Previously, flexibility was changing a system to fit the environment. A community gains flexibility by giving citizens freedom, i.e. the right to not be a slave[49]. It allows local resource control to increase performance, as do decentralized network protocols like Ethernet.

2. *Order*. Previously, reliability was a system's ability to survive internal part failure or error. A community gets reliability by order, that citizens, by rank, role or job, know and do their duty. Some cultures set up warrior or merchant castes to achieve this. Online order is also by roles, e.g. Sysop or Editor.

Freedom and order are in tension, as freedom has no class but order does. *Democracy* merges freedom and order, as free citizens select an order hierarchy, not just of President or Prime Minister, but for all positions. Democracy is rare online, but Slashdot uses it.

1. *Ownership*. Previously, security was a system's defense against outside takeover. A community is secure internally by ownership, e.g. to "own" a house guarantees that if another takes it, the community will step in[50]. Online, ownership works by access authorization.

48. The problem with competition is that if you give peanuts you get monkeys but if you give honey you get wasps, while with cooperation helping others doesn't help them help themselves.

49. Physical slavery is tyranny, informational slavery is propaganda and psychological slavery is political correctness, where the few tell the many how to think. Wanting to live other's lives is a sign of emptiness.

50. Note: state ownership, as in communism, is still ownership. Real ownership can be given away. When a state gives ownership away, one gets freedom.

2. *Openness*. Previously, extendibility was a system using what is outside itself. A community doing this was America's invitation to the world:

"Give me your tired, your poor, your huddled masses yearning to breathe free."

A society is open internally if any citizen can achieve any role by merit, as Abraham Lincoln, borne in a log cabin, became US president. The opposite is nepotism or cronyism, giving jobs to family or friends. If community advancement is by who you know not what you know performance reduces. Open source systems like Source Forge let people advance by merit.

Ownership and openness are in tension, as the right to keep out denies the right to go in. *Fairness* can reconcile public access and private control. Offline fairness is based on justice systems but online it is supported by code.

1. *Connectivity*. Previously, connectivity was the ability to open communication channels. Communities connect internally by media like TV, newspapers, radio, and now the Internet. A centrally controlled press is propaganda, while a free press lets everyone put a point of view.

2. *Privacy*. Previously, privacy was a citizen's right to control information about themselves. It is the ownership of self-data, not secrecy, so it includes the right to make personal data public.

Connectivity and privacy are in tension, as opening a channel to connect can reveal personal data. *Transparency* illustrates a combination, as public officials are entitled to privacy, except if acting for a community. Transparency is a citizens right to see governance on their behalf, including money spent and privileges given.

In summary, a community must increase citizen competence to be productive, increase trust and deny crime to get synergy, give freedoms to adapt and innovate, establish order to define responsibilities, allocate ownership to prevent property conflicts, be open to talent outside and inside[51], be connected to generate agreement, and grant privacy to relieve citizens from the pressure of social interaction. All these increase social performance and prosperity.

### 24.3.6 Synergy

Social synergy arises when people work to create *each other's* outcomes. It isn't just people adding efforts, say to lift a heavy log together. Positive synergy is the majority adding value to others and negative synergy is reducing it, e.g. war. Trade is mutual synergy, when my acts give your benefits, e.g. a fisherman who trades fish for a farmer's grain turns excess into value. Each gives an extra they don't really need for a deficit they lack. Modern prosperity arises when specialists share, and specialists produce nearly everything we use[52].

Synergy is even greater for information, as one can give information to others without losing it oneself. As connected communities grow and work at higher levels, they produce more but synergize *much more*. Productivity adds with size but synergy multiplies, because it depends on the number of interactions, not the number of citizens. Synergy is the key to prosperity in large connected societies (Wright, 2001)because it "expands the pie", making every slice larger. In contrast, zero-sum gains like war expand one slice at another's expense. Communities that generate synergy are "civilized".

Game theory, the formal calculation of personal gain and loss in social interactions, points out the fly in this social ointment. If my acts make your gain and yours make mine, what if I take from you and give nothing back? In fact, on the personal level, it always pays to *defect,* e.g. for a seller to give shoddy goods or for a buyer's check to bounce. But if the cheated "sucker" doesn't repeat the

51. Sexism and racism are community level losses. If women can't work, half the population can't add to its productivity. If a race, like black people, are excluded, so are their contributions.

52. How many of the objects you use each day do you make? How many are even produced in your country?

interaction, both lose their synergy gains, so cheaters destroy their own success. Synergy is destroyed by anti-social defections, or crime. Social dilemmas are common in society, e.g. social loafing, the volunteer dilemma and the tragedy of the commons[53]. The predicted equilibrium is that all parties defect (Poundstone, 1992), i.e. that synergy is unstable. The mystery isn't why people don't trust but why they do.

The answer proposed here is that people evolve a *community sense*, when it doesn't pay to defect, e.g. a community overgrazing its commons loses a valuable resource forever[54]. Social dilemmas can't be solved at the personal level, as an honest person among cheats is just a sucker. Only community level action changes the social unit and the gain-loss equation, as explained in detail elsewhere (Whitworth and Whitworth, 2010).

As people, we struggle to see social acts are hard on a community level. A theft that is "good" for a robber is "bad" for the victim, but for a community, theft is always bad. Why spend thousands of dollars in police, court and prison costs to prosecute a hundred dollar theft? For a community, it is a good deal, as crimes that succeed create copycats. The main reason people cheat is because "*everyone is doing it*" (Callahan, 2004), so one defection can snowball into a social collapse, i.e. no synergy[55]. Giuliani's clean up of crime in New York[56] cost millions but the community synergy gain was billions.

---

53. In social loafing people let the rest of the group do the work, e.g. people pull a rope less when with others than when alone. In the volunteer dilemma, people don't volunteer to help a group because someone else will. In the tragedy of the commons, farmers overgraze and destroy a commons because if they don't, others will.

54. The tragedy of the commons exemplifies whaling, forest and wildlife conservation issues. A community destroying a common resource is like a farmer killing all his cows to eat. It is stupidity not profit.

55. For example, if a fast-food restaurant is kept clean people drop less rubbish. If it is messy, they drop more.

56. By Wilson's 'Broken Windows Theory'

| Purpose | Examples | Synergy | Defection |
|---|---|---|---|
| *Communicate* | Email, Chat, ListServ, IM | *Shared communication*: People send more useful messages | *Spam*: Spammers waste others time, giving spam filters. |
| *Learn* | Moodle Blackboard | *Shared learning*: Students help others learn, reduce bottlenecks | *Plagiarism*: Student copying gives systems like Turnitin.com. |
| *Know* | Wikipedia, Tiddlywiki | *Shared knowledge*: Taps group knowledge, not just a few experts | *Trolls*: Wikipedia's monitors fight knowledge "trolls". |
| *Friend* | Facebook, Myspace | *Relationships*: People keep in touch with friends and family | *Predation*: Social networks report and banish predators |
| *Keep current* | Digg, Del.icio.us | *Shared bookmarks*: Social bookmarks let people see trends. | *Advocates*: Who "digg" a web site they own. |
| *Play* | Second Life, Sims | *Shared play*: Avatars experience things impossible in reality. | *Bullies/Thieves*: Newbies robbed by veterans need "safe" areas. |

| | | | |
|---|---|---|---|
| *Trade* | E-Bay, Craig's List, Amazon | *Item trading*: People from anywhere exchange more goods. | *Scams:* Scams are reduced by online reputation systems. |
| *Work* | Monster | *Work trading:* People find and offer work more easily. | *Faking*: Padded CVs and fake job offers need reputation systems. |
| *Download* | Webdonkey, BitTorrent | *Shared down-loading*: Groups share processing downloads. | *Piracy*: Prosecutions by society's copyright laws. |
| *Publish* | Flickr, YouTube | *Shared experience:* People share photos and videos. | *Offensiveness*: Editors remove items that offend. |
| *Advice* | Help boards AnandTech | *Technical advice:* People who have solved problems help others | *Confusers:* People who ask questions before checking old ones are scolded. |
| *Discuss* | Slashdot, Boing-Boing | *Shared views:* People comment and read others opinions easily | *Caviling*: Karma systems deselect those who just "peck" new ideas |
| *Follow* | Twitter | *Forms a group view* by linking leaders and followers. | *Identity theft*. A leader's online persona can be hijacked. |

Table 24.8: Socio-technical synergies and defections.

Socio-technical systems not only deny defections, but also enable synergies (Table 24.8). Forums like AnandTech illustrate this, as if anyone in a group solves a problem everyone can get the answer. The larger the group, the more likely someone can solve in seconds a problem you have struggled with for days. Same again functions let Amazon readers use the experiences of others to find books bought by those who bought the book they are looking at now. Wikipedia users correct errors of fact, supply references and examples to everyone.

> Synergy reduces when citizens work to personal requirements like:
>
> "*Take what you can and give nothing back*"
>
> Synergy increases when citizens follow community ethics like:
>
> "*Give unto others as you would they give unto you*".

Personal ethics is community pragmatics because without the former there is no social synergy, and without synergy there is no community prosperity. If synergy gains return to the people who generate them, the society will be stable. Previously, only heroes, of art, science, music, politics or other, gave to society. Today, socio-technology lets us all be "small heroes", giving back to a community that gives to us. The miracle of socio-technology is that people will help others for no personal gain whatsoever[57].

57. Social evolution require personal evolution. Social health is the percentage of citizens who hand in a lost wallet. Without social health, self-service supermarkets would fail, and without self-service we would queue for hours for goods. Online communities, selected by the digital divide, have more social health, and so predict our social future.

### 24.3.7 Communication performance

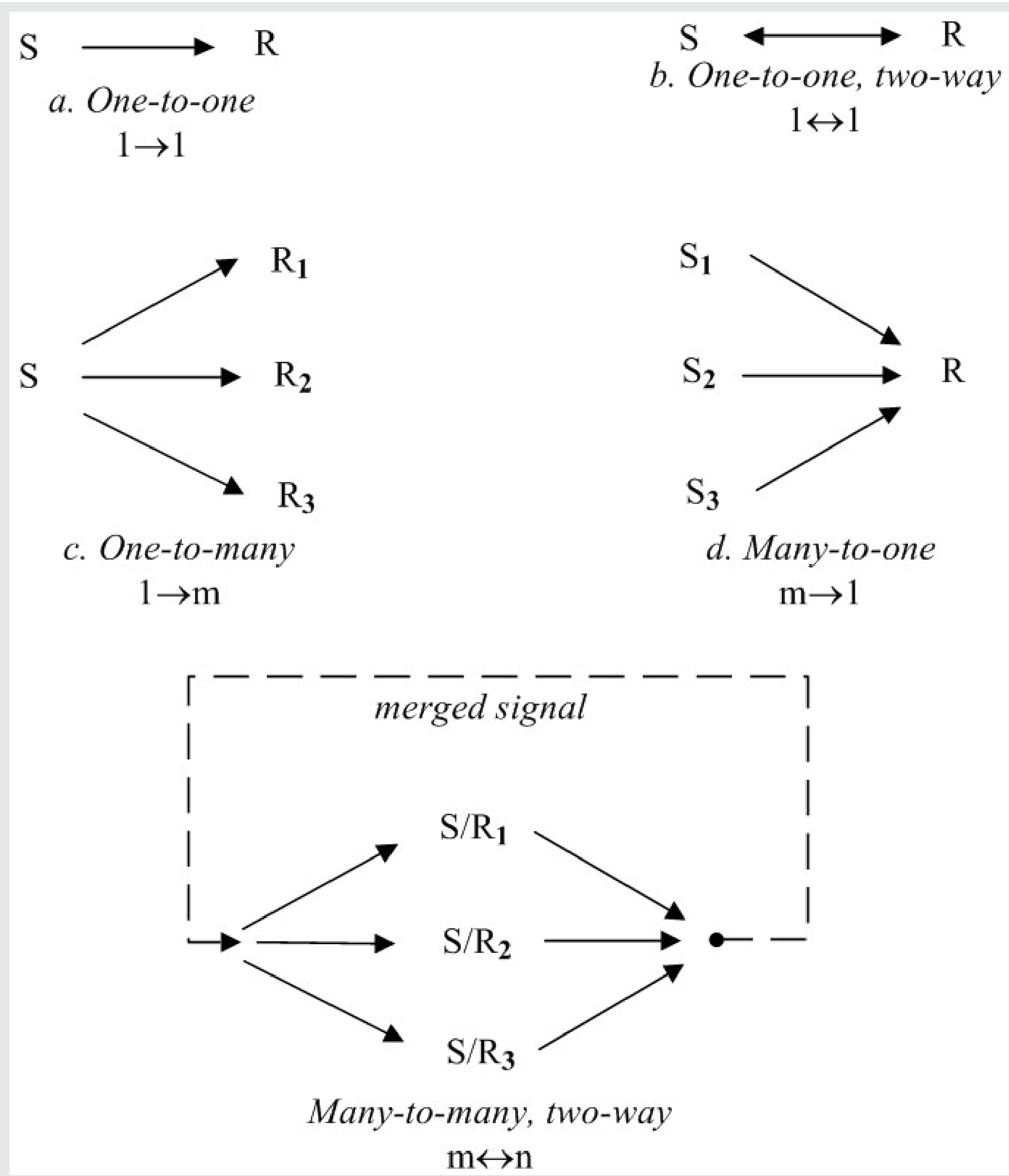



Figure 24.16: Linkage types (S = Sender, R = Receiver).

Communication media transmit meaning between senders and receivers. Meaning is any change in a person's thoughts, feelings or intents. *Communication performance* is then the total meaning exchanged by a transmission, i.e. its sum human impact.

*Richness*. Part of communication performance is richness, the amount of meaning a message conveys. To see video as automatically richer than text confuses meaning richness with information bandwidth. Meaning is the impact on a person, so texting "I'm safe" can have more meaning[58] than a multi-media marketing video. Hence video phones didn't immediately replace audio phones and lean texting is still used. Media richness can thus be classified by the symbols that generate meaning, as follows:

1. *Position*. A single, static symbol, e.g. to raise one's hand.
2. *Document*. Many static symbols that form a pattern with meaning, as words form a sentence by syntax or as pixelsform an objectby gestalt principles. Documents are text or pictures.
3. *Dynamic-media (Audio)*. A dynamic channel with multiple semantic streams, e.g. speech has tone of voice and content[59]. Music has melody, rhythm and timbre.
4. *Multi-media (Video)*. Many dynamic channels, e.g. video is audio and visual channels. Face-to-face communication uses many sensory channels.

One expects richer media to have the potential to transfer more meaning.

*Linkage*. The meaning exchanged also depends on the number of senders and receivers, i.e. on *linkage* (Figure 24.16), which can be:

58. One may be overwhelmed to hear a loved one is safe, but completely ignore a sales video.

59. A semantic stream is the meaning produced by human processing. A physical channel processed differently can have many semantic streams if, e.g. tone of voice and message content are different streams.

1. *Interpersonal* (one-to-one, two-way): Both parties can send and receive, usually signed.

2. *Broadcast* (one-to-many, one-way): From one sender to many receivers, can be unsigned.

3. *Matrix* (many-to-many, two-way): Many senders to many receivers, usually unsigned.

As people have interpersonal communication so communities communicate group-to-group by matrix communication. This most powerful linkage is when many send and many receive in one transmit operation. It combines one-to-many (broadcast) and many-to-one (merging) communication (Figure 24.16). Addressing an audience is one-to-many communication, applauding a speaker is many-to-one, and an applauding audience to itself is matrix communication. In the latter case, the group producing the clapping message also receives it. Matrix communication allows normative influence, so audiences can start and stop clapping together. A choir singing is matrix communication, so when choirs go off key, they usually do so together.

Face-to-face groups use matrix communication, as body language and facial expressions convey everyone's position on an issue. A valence index, calculated from member position indicators, can predict a group discussion outcome as well as the words (Hoffman and Maier, 1961). So online electronic groups can form social agreement using only anonymous, lean, many-to-many signals, with no rich information exchange or discussion (Whitworth et al, 2001). Community voting, as in an election, is a physically slow matrix communication that computers can speed up. Tag cloud, reputation system and social book-mark technologies all illustrate online support for matrix communication.

If communication performance is richness and linkage, a tyranny bombarding citizens 24/7 with TV video propaganda is low linkage (one-to-many) while people talking freely via text blogs is high linkage (many-to-many), i.e. the latter may communicate more.

### 24.3.8 Communication media

Table 24.9 shows various communication media by richness and linkage, with electronic forms in italics, e.g. a phone call is an interpersonal audio but a letter is interpersonal text. A book is a broadcast document, but radio is broadcast audio and TV is broadcast video. The Internet can broadcast documents (web sites), audio (podcasts) or videos (YouTube). Email allows two-way interpersonal text messages, while Skype adds two-way audio and video. Chat is few-to-few matrix text communication, as is instant messaging but with known people. Blogs are text broadcasts that also allow comment feedback. Online voting is matrix communication, as many communicate with many in one operation.

Computers allow "anytime[60], anywhere" communication for less effort, e.g. an email is easier to send than posting a letter. Lowering the message threshold means that more messages are sent (Reid et al, 1996). Email stores a message until the receiver can view it[61], but a face-to-face message is ephemeral, it disappears if you aren't there to get it. Yet being unable to edit the message sent makes sender state streams like tone of voice more *genuine*.

60. Asynchronous communication like email lets senders ignore distance and time but synchronous communication like Skype doesn't. One can't call someone who is asleep. So while the world is flat (Friedman), *the day is still round*.

61. For a physical network, asynchrony depends on the buffer capacity of its nodes.

| | Linkage | | |
|---|---|---|---|
| **Richness** | *Broadcast* | *Interpersonal* | *Matrix* |
| *Position* | Footprint, Flare, Scoreboard, Scream, | Posture, Gesture, Acknowledgement, Salute, *Smiley* | Show of hands, Applause, An election, *Web counter, Karma system, Tag cloud, Online vote, Reputation systems, Social bookmarks* |
| *Document* | Poster, Book, *Web site, Blog, Online photo, News feed, Online review, Instagram, Twitter*[1] | Letter, Note, *Email, Texting, Instant message, Social network*[2] | *Chat, Twitter*[1], *Wiki, E-market, Bulletin board, Comment system, Advice board, Social network*[2] |
| *Dynamic-media (Audio)* | Radio, Loud-speaker, Record or CD, *Podcast, Online music* | Telephone, Answer-phone, *Cell phone, Skype audio* | Choir, Radio talk-back, *Conference call, Skype conference call* |

| | | | |
|---|---|---|---|
| *Multi-media (Video)* | Speech,<br>Show,<br>Television,<br>Movie, DVD,<br>*YouTube video* | Face-to-face conversation,<br>*Chatroulette*<br>*Video-phone,*<br>*Skype video* | Face-to-face meeting,<br>Cocktail party,<br>*Video-conference,*<br>*MMORPG*<br>*Simulated world* |
| [1] Combines broadcast (text) and matrix (follow).<br>[2] Combines interpersonal and matrix. | | | |

Table 24.9: Communication media by richness and linkage.

Electronic communication was expected to just become richer, with video the anointed heir, but EBay's reputations, Amazon's book ratings, Slashdot's karma, tag clouds, social bookmarks and Twitter aren't rich at all. Table 24.9 shows that computer communication evolved by *linkage as well as richness*. Computer chat, blogs, messaging, tags, karma, reputations and wikis are all high linkage but low richness.

Communication that combines richness and linkage is interface expensive, e.g. a face-to-face meeting has rich channels and matrix communication to give sender state information and resolve real time contentions like people talking at once by showing where others are looking. To do this online requires many video streams on a screen, but who then controls the interface? Does each person control their own, and ignore the rest, or does one person set a common interface? In *audio-based tagging,* a person speaking automatically makes their video central (Figure 24.17). The interface is common but it is group-directed, i.e. democratic. *Gaze-based tagging* is the same except that when people look at a person their window expands, as when many people use a link it gets bigger. It is in effect a *group directed bifocal display* (Spence and Apperley, 2012). Only when matrix communication is combined with media richness will online meetings start to match face-to-face ones.







Figure 24.17 A-B: Audio based video tagging.

As video-phones are now easily available, why isn't video-phoning the norm? Perhaps it has disadvantages, like having to dress-up for a call or check the background before calling Mum. Some may prefer text to video precisely because it is less rich, if they *don't* want to communicate. Computer communication isn't just about richness because communication isn't just about the message — there is the sender and receiver too.

### 24.3.9 Semantic streams

Communication goals can be classified by level as follows (Whitworth et al, 2000):

1. *Informational*. The goal is to analyze information about the world and decide a best choice. This logical process is surprisingly fragile (Whitworth et al, 2000).

2. *Personal*. The goal is to form relationships which are more reliable. Relating involves aturn-taking, mutual-approach process, to manage the emotional arousal evoked by the presence of others (Short et al, 1976).[62]

3. *Community*. The goal is to stay "within" the group, as belonging to a community means being part of it, and so protected by it. Communities outlast friends.

62. The Haka communicates between two Maori warbands or tribes. It conveys an intent as well as a state, see here: http://www.youtube.com/watch?v=c-lrE2JcO44

| Goal | Influence | Link-age | Questions |
|---|---|---|---|
| *Analyze* (task information) | Informational influence, of the facts | Broad-cast | What is right? What is best? |
| *Relate* (to other people) | Personal influence, of other people | Inter-personal | Who do I like? Who do I trust? |
| *Belong* (to a community) | Normative influence, of the community | Matrix | What is everyone doing? Am I "in" the group? |

Table 24.10: Human goals by influence and linkage.

Table 24.10 shows how each goal maps to influence and linkage. Whether online or off, we *analyze* information, *relate* to others and *belong* to communities, so are subject to informational, personal and normative influence. The latter is based neither on logic nor friendship, e.g. patriotism is my country right or wrong, friendships or not. An individual may be influenced by task information, friend recommendations or community norms via different semantic streams. Semantic streams are people processing a physical signal in different ways to generate different meanings, where *one* physical message can *at the same time* convey:

1. *Message content*. Symbolic statements about the literal world, e.g. a sentence.
2. *Sender state*. Sender psychological state, e.g. an agitated tone of voice.
3. *Group position*. Sender intent over many is a group intent, e.g. an election.

Human communication is subtle becaus77e one message can have multiple meanings and people respond to many semantic streams at once, e.g. a person leaving a party may *say* "I had a good time", but by *tone* imply the opposite. One can say

'I AM NOT ANGRY!" in an angry voice[63]. What is less obvious is that a message can also indicate a *position,* or intent to act, e.g. saying "I had a good time" in a certain tone or body language can indicate an intention to leave a party. When a community acts, its citizens follow.

In the general model (Figure 24.18), physical level signals generate many semantic streams and influences. While face-to-face interactions allow multi-stream communication, computing tends to pick one type, e.g. email text gives content but not sender state. Online voting gives position but not comments. Technologies that operate at the community level use matrix or group-to-group communication, such as:

a. The reputation ratings of Amazon and E-Bay are community-based product quality control and Slashdot does the same for content, letting readers rate comments so viewers can filter out low quality ones.

b. Social bookmarks, like Digg and Stumbleupon, let users share link favorites, to see what the community is looking at.

c. Tags are technology to increase the font size of links according to frequency of use. As peopl]e walk in forests on the paths trod by others, so we can follow the "web-tracks" of others on a browser screen.

d. Twitter's follow function lets people see the leaders they like, and lets leaders broadcast ideas to followers.

The power of the computer is to allow matrix communication by millions and billions. What might a global referendum on current issues reveal? The Internet could tell us.

As for the future, in an Internet dominated by personal "apps", multi-user apps are an obvious next step, as are applications supporting many semantic streams, like Facebook friend voting. Given recent advances in connectivity, we

63. Some people may not process the sender state semantic stream, e.g. those with autism.

can expect a "bite-back" in privacy demands, i.e. more small groups or "tight" communities that are harder to get in.

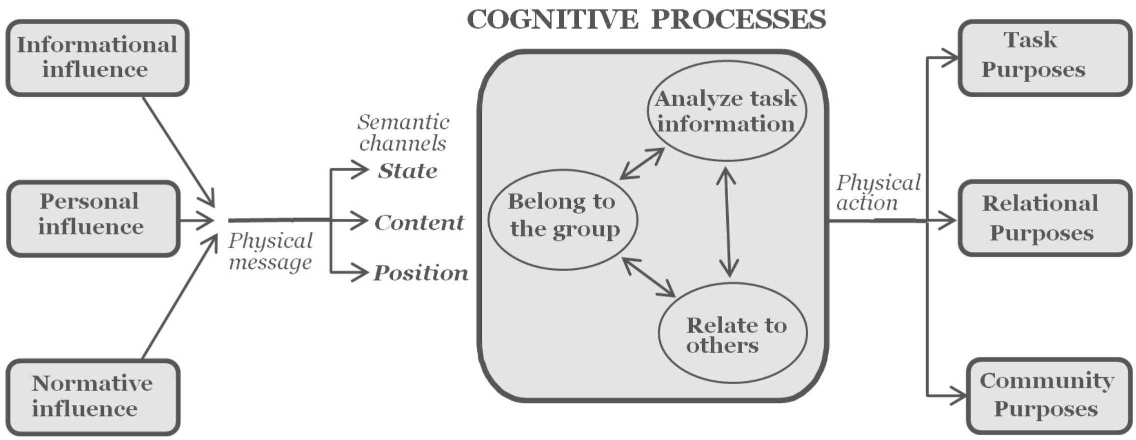



Figure 24.18: Cognitive processes in communication.



The World Wide Web is a system evolving. Its first level, an information library accessed by search tools, is well in place. The second, a medium for personal relations, is also well underway. The third, a civilized social environment, is the current and future challenge. Even a cursory study of Robert's Rules of Order will dispel any illusion that social dealings are simple (Robert, 1993). Socio-technology allows hundreds of millions of people to act together but we are still working out what "here comes everybody" (Robert, 1993). Socio-technology allows hundreds of millions of people to act together, but we are still figuring out what "Here comes Everybody"[64] means (Shirky, 2008). None of us is an island, as we link to mothers, fathers, brothers, sons, daughters, aunts, sisters, wives, grandmothers, uncles, grandfathers husbands and friends, so when social others talk, even hardened dictators listen. This is a good thing.

64. Question: 'Where does an 800lb Gorilla sit when it comes to dinner?' Answer: Anywhere it wants to. Communities are like this, but to act they must agree, which can take months, years or decades.

### 24.3.10 Discussion questions

Research selected questions from the list below. If you are reading this chapter as part of a class - either at university or a commercial course - you can research these questions in pairs and report back to the class, with reasons and examples.

1. Why can't technologists leave the social and ethical questions to non-technologists? Give examples of IT both helping and hurting humanity. What will decide, in the end, whether IT helps or hurts us overall?

2. Compare central vs. distributed networks (Ethernet vs. Polling). Compare the advantages and disadvantages of centralizing vs. distributing control. Is central control ever better? Now consider social systems. Of the traditional socio-technical principles listed, which ones distribute work-place control? Compare the advantages and disadvantages of centralizing vs. distributing control in a social system. Compare governance by a dictator tyrant, a benevolent dictator and a democracy. Which type are most online communities? How might that change?

3. Originally, socio-technical ideas applied social requirements to work-place management. How has it evolved today? Why is it important to apply social requirements to IT design? Give examples.

4. Illustrate system designs that apply: Mechanical requirements to hardware[65]. Informational requirements to hardware[66]. Informational requirements to software. Personal requirements to hardware[67]. Personal requirements to software. Personal requirements to people. Community requirements to hardware. Community requirements to software. Community requirements to people. Community requirements to communities. Give an example in each case. Why not design software to mechanical requirements?

65. As specified by engineering

66. As specified by computer science.

67. As specified by psychology.

5. Is technology the sole basis of modern prosperity? If people suddenly stopped trusting each other, would wealth continue? Use the 2009 credit meltdown to illustrate your answer. Can technology solve social problems like mistrust? How can social problems be solved? Can technology help?

6. Should an online system gather all the data it can during registration? Give two good reasons not to gather or store non-essential personal data. Evaluate three online registration examples.

7. Spam demonstrates a socio-technical gap, between what people want and what technology does. How do users respond to it? In the "spam wars", who wins? Who loses? Give three other examples of a socio-technical gap. Of the twenty most popular third-party software downloads, which relate to a socio-technical gap?

8. What is a legitimate government? What is a legitimate interaction? How do people react to an illegitimate government or interaction? How are legitimacy requirements met in physical society? Why won't this work online? What will work?

9. What is the problem with "social engineering"? How about "mental engineering" (brainwashing)? Why do these terms have negative connotations? Is education brainwashing? Why not? Explain the implications of all this for STS design.

10. For a well known STS, explain how it supports, or not, the eight proposed aspects of community performance, with screenshot examples. If it doesn't support an aspect, suggest why. How could it?

11. Can one own something but still let others use it? Can a community be both free and ordered? Can people compete and cooperate at the same time? Give a physical and online examples. How are such tensions re-

solved? How does democracy reconcile freedom and order? Give examples in politics, business and online.

12. What is community openness for a nation? For an organization? For a club or group? Online? Why are organizations that promote based on merit more open? Illustrate technology support for merit-based promotion in an online community.

13. Is a person sending money to a personal friend online entitled to keep it private? What if the sender is a public servant? What if it is public money? Is a person receiving money from a personal friend online entitled to keep it private? What if they are a public servant?

14. What is social synergy? What destroys it? How do communities encourage synergy? How do they prevent its destruction? How do trust and synergy relate? Give physical and electronic examples.

15. Give five examples of defections in ordinary life. What happens if everyone defects? Give five online examples, and for two specify how technology lowers defections.

16. Would you prefer to be a middle class citizen now or a lord three hundred years ago? Consider factors like diet, health, clothes, leisure, travel, etc. Where did the lord's wealth mainly come from? Where does the power of your salary to buy many things come from today? What is the principle and how does it apply online?

17. What is a social dilemma? Give three physical examples from your experience. Why can't individuals solve them? How are they solved? Give three online social dilemmas. How are they to be solved? Relate this to socio-technical design.

18. What happens if no-one in a group suggests anything? What happens if you suggest things in a group? How can groups manage this? Answer the same questions for volunteering. Give examples from your experience. What percentage of online users are "lurkers", who look but don't post? Review a popular board you haven't used before. What stops you contributing? Add something anyway. How could the board increase participation?

19. Is ethics idealism or pragmatism? Explain the statement: Personal ethics is community pragmatics. Consider a thief who steals a wallet and isn't caught. List the thief's gains and the victim's losses. What is the net community result? What happens if everyone in a community steals, i.e. takes but does not give? Generalize to online cases. How then does STS design relate to ethics?

20. Why is synergy more important for larger communities? Why is it especially important for socio-technical systems? How can technology help increase synergy? Report the current estimated sizes of popular socio-technical systems. Clarify what is exchanged, who interacts and the synergy.

21. What is communication? What is meaning? What is communication performance? How can media richness be classified? Is a message itself rich? Does video always convey more meaning than text? Can rich media deliver more communication performance? Give online and offline examples.

22. What affects communication performance besides richness? How is it classified? Is it a message property? How does it communicate more? Give online/offline examples.

23. If media richness and linkage both increase communication power, why not have both? Describe a physical world situation that does this? What is the main restriction? Can online media do this? What is, currently, the main contribution of computing to communication power? Give examples.

24. What communication media type best suits these goals: telling everyone about your new product; relating to friends; getting group agreement? Give online and offline examples. For each goal, what media richness, linkage and anonymity do you recommend. You lead agile programming team spread across the world: what communication technology would you use?

25. State differences between the following media pairs: email and chat; instant messaging and texting; telephone and email; chat and face-to-face conversation; podcast and video; DVD and TV movie; wiki and bulletin board. Do another pair of your choice.

26. How can a physical message convey content, state and position semantic streams? Give examples of communications that convey: content and state; content and position; state and position; and content, state and position. Give examples of people trying to add an ignored semantic stream to technical communication, e.g. people introducing sender state data into lean text media like email.

27. Can a physical message generate many information streams? Can an information stream generate many semantic streams? Give examples. Does the same apply online? Use how astronomical or earthquake data is shared online to illustrate your answer.

28. You want to buy a new cell-phone and an expert web review suggests model A based on factors like cost and performance. Your friend recommends B, uses it every day, and finds it great. On an online customer feedback site, some people report problems with A and B, but most users of C like it. What are the pluses and minuses of each influence? Which advice would you probably follow? Ask three friends what they would do.

29. What is the best linkage to send a message to many others online? What is the best linkage to make or keep friends online? What is the best link-

age to keep up with community trends online? List the advantages and disadvantages of each style. How can technology support each of the above?

30. Explain why reputation ratings, social bookmarks and tagging are all matrix communication. In each case, describe the senders, the message, and the receivers. What is the social goal of matrix communication? How exactly does technology support it?

31. Give three online leaders searched by Google or followed on Twitter. Why do people follow leaders? How can leaders get people to follow them? How does technology help? If the people are already following a set of leaders, how can new leaders arise? If people are currently following a set of ideas, how can new ideas arise? Describe the innovation adoption model. Explain how it applies to "viral" videos?

## 24.4 PART 4: AN EXAMPLE: ONLINE RIGHTS

**"**A right is a community permission to act**"**

Legitimacy analysis specifies community requirements for technology design. Previous examples are polite computing (Whitworth and Liu, 2008) and channel email (Whitworth and Liu, 2009). This section proposes an access control model based on these social requirements:

A. *Ownership*. To reduce object conflicts.

B. *Freedom*. To own oneself, to not be a slave.

C. *Fairness*. That social consequences reflect action contributions (Rawls, 2001).[68]

68. It is unfair to B if A's acts cause B's loss only, and unfair to A if A's acts cause B's gain only.

D. *Privacy*. To control the release of personal information to others.

E. *Transparency*. A democratic citizen's right to know how they are governed.

If the new user of computing is society, we must specify its requirements.

### 24.4.1 Access control

In computing, decision support systems recommend decisions, access control systems permit them and control systems carry them out. Access control began with multi-user computing as users sharing the same system came into conflict (Karp et al., 2009). Traditional access control systems (ACSs) use a subject by object access permission matrix to allocate rights (Lampson, 1969). As computing evolved, ACS logic offered local access control for distributed systems and roles for many person systems. With these variants, the matrix approach has worked for military (Department of Defense, 1985), commercial (Clark and Wilson, 1987), organizational (Ferraiolo and Kuhn, 2004), distributed (Freudenthal et al, 2002), peer-to-peer (Cohen, 2003) and grid environment (Thompson et al, 1999) applications.

Today, access control in *social networks* (SNs) is more about access than control. The permission matrix for friend interactions increases geometrically, not linearly, with group size, so for hundreds of millions of people the possible connections are astronomical. Each account also adds hundreds or thousands of photos or comments a year. Finally, each person wants the sort of domain control previously reserved only for system administrators. Social networkers want *local access control*, not just to read, write and execute files (Ahmad and Whitworth, 2011), but to control their own social structure, without asking a central authority for permission (Sanders and McCormick, 1993),e.g. to restrict a photo to family or friends. Social networks vastly increase ACS complexity, as millions of users want all rights to billions of resources, plus rights to re-allocate rights. They are the perfect storm for the traditional ship of access control.

The current rules of social network interaction are based on designer intuitions rather than formal models, so they vary between systems and over time, with public outrage the only check. There is no agreed scheme for allocating permissions to create, edit, delete or view object entities, let alone manage roles. The aim here is to fill that gap, to develop a socio-technical access control model that is legitimate, efficient, consistent and understandable.

### 24.4.2 Rights

Communities, by norms, laws or culture, grant citizens *rights,* or social permissions to act. Rights reduced physical conflict, as parties who agree on rights don't have to fight. This moved the conflict from the physical level to the informational or legal level[69]. Physical society expresses rights in terms of ownership (Freeden, 1991), so specifying who owns what online can specify rights in a way that designers can support and users can understand (Rose, 2000). This doesn't mechanize online interaction, as rights are choices not obligations, e.g. the right to sue doesn't force one to sue. Legitimate access control defines what online actors *can* do not what they *must* do.

Traditional design refers to software "users", as if they were on a drug, but Facebook's users aren't part of the software. Socio-technology talks of *actors* who switch software, not passive users. As shops can see "a sale" or "a customer", IT designers see a user or an actor.

An actor is a system able to act independently of outer conditions, i.e. to act not react. Actors can initiate acts, which implies some internal choice or autonomy[70]. A program that always responds the same way to the same input has no autonomy, so can't itself be an actor[71].

69. Personal acts between people is the level after that, when people drop rules and believe in each other.

70. From the Greek autos 'self' and nomos 'law,' i.e. a system that can make its own laws. It is not all or none, e.g. poke a ball with a stick and it moves, poke a dog and it runs away or bites you, poke a man and he might do any of the above, or take the stick off you.

71. It can however be an agent.

A person is an actor with an ego-self and a citizen is a person who can be held to account[72]. To hold to account, to link consequences to people, is fundamental to all social interaction[73]. By accountability, communities reward those who benefit it and punish those who harm it.[74] While philosophers argue over free will, all communities consider citizens accountable and govern accordingly. Those deemed not so, the criminal or insane, are in the care of those who are. A community holds citizens[75] to account for the effects of their acts not on themselves *and* on others. Accountability is the over-arching social requirement, without which communities fail. It only applies to people, e.g. in car accidents the driver is held to account not the car, as the car has no personal self to be accountable.[76]

Rights arise when social requirements manifest as personal cognitions, which manifest as informational rules, which manifest as action directives. In physical communities, police and courts direct citizens to follow laws, written by judges who understand justice. Online, the same applies, but in this architecture code is the law, police, judge, jury and prison guard. To not be corrupt, systems must be legitimate by design.

The following derives informational rights from community requirements stated on the personal level. In information terms, a right is an actor (A) applying an operation (O) to an entity (E):

*Right = (Actor, Entity, Operation) = (A, E, O)*

72. Accountability only assumes some choice at some point, e.g. a drunk with no control can be fined if he earlier chose to drink too much. Did a drug addict who can't stop now, but once could, choose that path? To argue no denies accountability, so a community can take control of their life anyway.

73. Community justice and law began with revenge, where people personally held others to account.

74. For example, laws punish those who steal and copyright rewards those who create.

75. A citizen is a person who in a community. A foreign visitor is not a citizen but still a person. A person is anyone who is accountable. A criminal who is not accountable is locked up in jail.

76. A company, as an informational entity, can't be accountable as it has no ego-self. To punish a cheating company by declaring it bankrupt lets its owners start another company to do the same thing again. Treating companies as people in the law was a great ethical and legal error of the last century. It underlies most of the scams of the wealthy.

Rights can be stored as (Actor, Entity, Operation)triplets, where an actor is an accountable entity or their agent, an entity is any object, actor[77] or right, and an operation is any one available to the entity. A right transmitted or stored is often called a permission.

### 24.4.3 Specification

Socio-technical systems can be modeled as *entities* and *operations*:

1. **Entities**. Stored as static information, with properties.
   a. *Actor*. An entity that can participate in a social interaction.[78]
      i. *Persona*. Represents an accountable offline person or group.
      ii. *Group*. A set of personae acting as one.[79]
      iii. *Agent*. An actor that represents another actor.
   b. *Object*. Conveys information and meaning.
      i. *Item*. A simple object with no dependents, e.g. a bulletin board post.
      ii. *Space*. A complex object with dependents, e.g. a bulletin board thread.
   c. *Right*. A system permission for an actor to operate on an entity.
      i. *Simple rights*. Rights to act on object or actor entities.
      ii. *Meta-rights*. Rights to act on right entities, e.g. delegate.
      iii. *Role*. A variable right (a set of rights).

77. An actor, being an entity, can act on itself. A persona can even delete itself, as a person can commit suicide.

78. A social actor need not be a person, e.g. a program can be an agent.

79. Online and offline are different worlds by their base architecture - the online world has information base. An offline group is physical people who act as one. Groups can also form groups, e.g. the stock market is a group of groups (companies). An offline group can have one online persona, e.g. a company registered on Facebook. An online group is a set of personae that act as one, so the access control system must define how it does this, see Sedtion 4.13.

2. **Operations**. Stored as a program or method that processes entities.
    a. *Null operations* don't change the target entity, e.g. view[80], enter.
    b. *Use operations* change the target in some way, e.g. edit, create.
    c. *Communication operations* transfer data from sender(s) to receiver(s), e.g. send.
    d. *Social operations* change a right or role, e.g. delegate.

Link operations are discussed elsewhere (Whitworth and Bieber, 2002).

### 24.4.4 The system itself

The information system itself is the first entity, owned by the system administrator (SA), who is the first user. A tyrant SA might alter posts or votes by whim but a benevolent dictator, Plato's best form of rule, gives citizens rights. As even benevolent dictators die, humanity invented democracy,to reduce dynasty transfer battles[81]. Yet no online system we know of votes for its system administrator, e.g. even Wikipedia isn't a democracy.

An ACS controls at the informational level. If it is *not* to be in charge, it must allocate all use rights to people who are accountable, giving the ACS operational principle:

P1. All non-null entity rights should be allocated to actors at all times.

So every only entity should be owned, ultimately, by a person.If this is not true, an access control system must at some point respond to an access request from itself. Yet as an information system, it has no self to act socially. Hence rights aren't added or deleted, but allocated and re-allocated.

80. View is null at the informational level but not at the psychological level, see later.

81. Compare the peaceful power transitions of democracies to the violence of dictatorial change.

## 24.4.5 Persona

An online persona represents an offline party, e.g. an avatar, profile, mail account, wall or channel can represent an offline person, group or organization. An online persona is activated by a logon operation, which equates it to the offline party. An online computer agent can act for a group, like installation software for a company, but social acts must ultimately trace back to people and online is no different[82]. If an installation misleads, we sue company directors not software[83].

*Who owns a persona?* Open systems let people self-register, to create their personae. If freedom applies online, one should own one's online self, but some systems don't permit this. Can you delete a Wikipedia or Wordpress profile?[84] The freedom requirement gives the ACS principle:

P2. A persona should be owned by itself.

Some complexities are that a persona can be:

*Abandoned*. HotMail accounts inactive for over 90 days are permanently deleted, i.e. if not used they "starve and die."

*Transferred*. One can permanently pass a persona to another, along with its reputation.[85]

*Delegated*. One can ask an agent act on one's behalf, e.g. a proxy vote.

*Orphaned*. If the person behind a persona dies, their will is physically respected, but online programs act as if death doesn't exist, e.g. one can get an eerie Facebook message from a person the day after going to his funeral. As in a few decades Facebook will represent millions of obituaries, we need online wills.

---

82. Registering by a nickname online instead of one's 'real' name denies accountability offline but not online, e.g. a banned EBay seller name loses its online reputation.

83. A person who acts as an agent can still be held accountable, e.g. if told to shoot someone and does do.

84. See how to permanently delete you account on popular web sites here: http://www.smashingmagazine.com/2010/06/11/how-to-permanently-delete-your-account-on-popular-websites/

85. In the movie The Princess Bride, the Dread Pirate Robert persona was passed on, so the idea is not new.

TABLE 24.11 BELOW SHOWS A SUMMARY OF PERSONA ACCESS RIGHTS

| **Persona** | *View* | *Delete* | *Edit* | *Ban* | *Create* |
|---|---|---|---|---|---|
| *SA* | √ | | | √ | √ |
| *Owner* | √ | √ | √ | | √[1] |

TABLE 24.11: Persona access rights. [1] *Delegated by the SA*..

## 24.4.6 Object entities

Object entities convey meaning by evoking cognitive processing, e.g. a family photo.

*Items*. A simple object with no dependents, e.g. a board post. It can be deleted, edited or viewed. In the object hierarchy tree, items are like leaves. An item can be a:

1. *Comment*: Items whose meaning depends on another, e.g. "*I agree*" makes no sense alone.
2. *Message*: Items with sender(s) and receiver(s), e.g. an email.
3. *Vote:* Items that convey a position, a choice from a response set.

*Spaces*. As leaves need branches, so items need spaces, e.g. an online wall that accepts photos is an information space - a complex object with dependents. It can be deleted, edited or viewed like an item, but can also contain objects, e.g. a bulletin board. Spaces within spaces give object hierarchies, with the system itself the first space.

A space is a *parent* to the *child* entities it contains, who depend on it to exist. So deleting a space deletes its contents, e.g. deleting a board deletes its posts. The move operation changes the parent space of an object. The enter space operation shows the objects on display in it. As every entity is in the system space:

P3: Every entity has a parent space, up to the system space.

If every entity has a parent space[86], its *ancestors* are the set of all spaces that contain it, up to the system itself, the first ancestor. The *offspring* of a space are any child objects it contains, their children, etc. So all entities have owners and ancestors, and any space can have offspring.

## 24.4.7 Operations

| *Entity Type* | *Operations* |
|---|---|
| Any entity | View |
| 1. Social entity | ..., Delete, Edit |
| a. Persona | ..., Logon |
| b. Agent | ..., Delegate |
| c. Group | ..., Join |
| 2. Object entity | ..., Delete, Edit, Move |
| a. Item | ..., ConvertToSpace |
| b. Space | ..., Create, Enter |
| 3. Right entity | ..., Allocate, Re-allocate |
| a. Role | ..., Friend, Ban |

Table 24.12: Operation sets by entity type.

86. Except, of course, for the system itself.

Operations are actor initiated methods on information entities subject to access control.

*Operation sets*. Operations can be clustered for access control purposes, e.g. delete flags an entity for destruction, undelete reverses that, and destroy kills it permanently. An ACS that can manage one can manage all. Likewise, edit alters entity values, append extends them, version edits with backup, and Wikipedia's revert is the inverse. Again, variants of a set present the same ACS issues, so to resolve one is to resolve all.

*Create*. While edit changes existing entity values, create adds a new entity, e.g. creating a Wikipedia stubfor others to edit. Duplicate is a variant of create. Table 24.12 shows the operation sets for various entity types, where create is an act on a space - see Section 13.10.

*View*. Operations like view are null acts that don't change their informational level target but viewing another is a personal level act. In *social facilitation,* knowing one is being looked at energizes the viewed party (Geen and Gange, 1983). Viewing someone affects them because success in a social group depends very much on how others see you. Privacy, to control information about ourselves, is important for the same reason. The act of viewing can have great effect on the community level, e.g. a "viral" online video makes others want to view it too.

The right use an entity implies accountability, but as one can't use what one can't see, use rights imply view rights, giving the ACS operational principle:

P4: Any right to use an object implies a right to view it.

*Communication*. In a simple communicative act, a sender creates a message that a receiver views. It is by definition a joint act where both parties have choice. Hence communication should be by mutual consent. Privacy is the right to remain silent, to not communicate and to not receive messages. In the physical world, people say "Can I talk to you?" because communication is by permission. Some online systems however, like email, don't recognize this. They give anyone the right to send a message to anyone, whether they will or no, and so invite spam. In contrast, in Facebook, chat, Skype and Twitter, one needs prior permission to message someone. The details of legitimate communication, where a *channel* is opened by mutual consent before messages are sent, are given in (Whitworth and Liu, 2009). The resulting ACS operational principle is:

P5: Any communication act should have prior mutual consent.

The evolution of telephony illustrates a communication evolution. At first phones just transmitted information — the phone rang and one answered, not knowing who was calling. This allowed telemarketing, the forerunner of spam. Now cell phones show caller id by default, so one can choose to respond, i.e. it is more mutual. Yet we still have to personally type in contact list names, while social networks synergize - we each type in our on name then let others add it to their contact list. Cell phone companies could use this synergy but like the makers of TV remotes, are locked into a one-level mind-set[87].

### 24.4.8 Roles

Roles, like parent, friend or boss, simplify rights management by covering many cases, but still remain understandable, so people can review, evaluate and accept them. They are equally useful online, e.g. Wikipedia citizens can aspire to steward, bureaucrat or sysop roles by good acts. Slashdot's automated rating system offers readers the moderator role (Benkler, 2002) if registered (not anonymous), regular users (for a time) with positive "karma" (how others rate their comments). Every registered reader has five influence points to spend on others as desired over a three day period (or they expire). In this role democracy, high rated commenter's get more karma points and so more say on who is seen. The technology lets a community democratically direct its governance.

87. Users would have the *option* to show a name instead of a number when they call. It could be their real name or a nickname, just as they now can choose to show 'Anonymous' instead of a caller-id number. The social system would self-adjust, if receivers chose not to reply to anonymous senders. If people chose to show their real name to the friends they call, that is their choice, so no privacy is lost. Privacy is not secrecy.

In information terms, a role is a variable rights statement, e.g. a friend role is a set of people with extra permissions. Roles are generic rights, giving the ACS operational principle:

P6: A role is a right expressed in general terms, as a pointer or set.

Roles are the variables of social logic:

*Role* = *(***Actor***,* **Entity***,* **Operation***)*

The bolding indicates a variable, e.g. the owner role can be generally defined as any party who has all rights to an entity:

$Role_{Owner} = (\mathbf{Owner}\ , Entity_{i}\ , Operation_{All})$

Making a person the owner just allocates the **Owner** pointer to their persona. Roles are flexible, e.g. the friend role lets one change who can see photos posted on a wall:

$Role_{Friend} = (\mathbf{Friend}\ , Entity_{Wall}\ , Operation_{View})$

where **Friend** is a persona set. To "friend" another is to add them to this role set, and to unfriend is to remove them. As a variable can be undefined, so a role can be empty, i.e. a null friend set. To "friend" is spoken of as act on a person, but it doesn't change the persona entity, so is really an act upon a local role. You decide your friends so don't need permission to friend anyone. Equally to ban a person adds them to the denied entry role for your space. If banning were an act on another's persona it, would need their consent. That it is an act on *my* role gives the ACS principle:

P7. A space owner can ban or give entry to a persona without its owner's permission.

Re-allocating actors isn't the only way to alter a role. By definition, one can change a role's:

1. *Actor*. The role actor set.

2. *Entity*. The entities it applies to.

3. *Operation*. The operations it allows.

For example, a friend role could limit the objects it applies to, with some photos for family only. It could also allow adding comments to photos or not. Few current systems fully use the power of local roles, e.g. social networks could let actors define an acquaintance role, with fewer rights than a friend but more than the public, or an extended family role.

### 24.4.9 Meta-rights

Owning an object is the right to use it:

$\text{RightUser} = \text{R}\,(\textbf{User}, \text{Entity}_i, \text{Operation}_{\text{Use}})$ ,

but a right as an entity can also be acted on, i.e. re-allocated. A meta-right is the right to re-allocate a right. In formal terms:

$\text{Right}_{\text{MetaRight}} = \text{R}\,(\textbf{Owner}, \text{Right}_{\text{Own}}, \text{Operation}_{\text{Allocate}})$ ,

where the entity acted on is a right. An owner with *all* rights to an entity also has its meta-rights, i.e. the right to change its rights. Paradoxically, fully owning an entity implies the right to give it away entirely. Reachability[88] requires meta-rights to be absolute, i.e. there are no meta-meta-rights. The gives the ACS operational principle:

P8. A meta-right is the right to allocate any entity right, including itself.

Previously to own an entity was to have all rights, but giving away use rights while keeping meta rights is still ownership, e.g. renting an apartment gives a tenant use rights, but the landlord still owns it, as they keep the meta-rights. The tenant can use it but the owner says who can use it.

88. Reachability, or halting, is that a program logic finishes and doesn't run endlessly.

## 24.4.10 The act of creation

To create an object from nothing is as impossible in an information space as it is in a physical one. Creation *cannot* be an act upon the object created, which by definition doesn't exist before it is created. An actor can't request ACS permission to create an object that doesn't exist. To create an information object, its data structure must be known, i.e. exist within the system. So creation is an act upon the system, or in general, an act on the space immediately containing the created object, giving the ACS operational principle:

P9. Creation is always an act on a space, up to the system space.

This rule is well defined if the system itself is the first space. Creating is an act upon a space because it changes the space that contains the created object. If creation is an act upon a space, the right to create in a space belongs to initially the space owner:

$Right_{Create} = R\ (SpaceOwner_i\ , Space_i\ , Operation_{Create})$

The right to create in a space initially belongs to its owner, who can delegate it to others. The logic generalizes well, e.g. to add a board post, YouTube video or blog comment requires the board, video, or blog owner's permission. One can only create in a space if its owner permits. Now an ACS can be simply initialized as a system administrator owning the system space with all rights, including create rights. The SA must then give rights away for a community to evolve. If the SA only delegates rights, they can always be taken back.

*Creator ownership*. Object creation is a simple technical act, but a complex social one, e.g. how are newly created entity rights allocated? The 17th Century British philosopher Locke argued that creators owning what they create is fair and increases prosperity, whether a farmer's crop, a painter's painting or a hunter's catch (Locke, 1963). If the creator of something chooses to sell or give it away, that is another matter. A community that grants producers the right to their products encourages creativity. Conversely, why produce for others to own? This gives the ACS operational principle:

P10. The creator of new entity should immediately gain all rights to it.

Creator ownership conveniently resolves the issue of how to allocate new object rights — they go to its creator, including meta-rights. This isn't what must happen, as a program can act any way it likes, e.g. to give all created object ownership to the system administrator. Creator ownership is a social requirement not a technical one, i.e. a condition of social success not a logical necessity. Such conditions can however be socio-technical axioms.

*Creation conditions*. Are when a space owner *partially* delegates creation, limiting:

1. *Object type*. The object type created, e.g. the right to create a conference paper isn›t the right to create a mini-track space.

2. *Operations*. The operations allowed on created objects, e.g. blog comments aren't usually editable once added, but ArXiv lets authors edit publications as new versions.

3. *Access*. Who can access created objects, e.g. YouTube gives contributors exclusive edit rights, but Wikipedia lets anyone edit any creation.

4. *Viewing*. Who can view created objects, e.g. bulletin boards let you see what others submit, but conferences in the paper review phase don't.

*Editing*. The field values of a created object, e.g. date added may be non-editable. The space owner may also set field default values.

A space owner can delegate creation rights as needed, e.g. to set vote results to only show to people who have voted, to avoid bias.

*Transparency*. Yet fairness dictates a creator's right to know creation conditions in advance. In general, transparency is the right to view rights that affect you. So those who create in a space should know the creation rules in advance. The ACS principle is:

P11. A person can view in advance any rights that could apply to them.

Successful socio-technical systems like Facebook, YouTube and Wikipedia do this.

In sum, a space owner can delegate the right to create in whole or part, but must disclose creation conditions up front so potential creators can decide if creation is worth it.

### 24.4.11 Role allocations

When an entity is created in a space, the system can assign the following roles:

- *Owner*. Has meta rights to the entity.
- *Parent*. The containing space owner.
- *Ancestor*. Ancestor space owners, with the SA the first ancestor.
- *Offspring*. The owners of any entities contained in a space.
- *Local public (space only)*. Actors who are permitted to enter the space.

A space owner owns its local public role can define what others can do or see in the space:

$$Role_{LocalPublic} = (\textbf{LocalPublic}\ , Space_i\ , Operation_{Any}\ )$$

It can be set manually, as friends are allocated, or point to a GlobalPubliclist.

*Ancestor role*. A conference paper's ancestors are its mini-track, track and conference chairs. An entity, being part of the space it exists in, must be visible to the owner of that space. Privacy doesn't contradict this, as it refers to the display of personal information not created object information. Generalizing, the ACS principle is:

P12. A space owner should have the right to view any offspring.

So the ancestor role for any entity is given view rights to it:

$Role_{Ancestor}$ = (**Ancestors**, *Entity*$_i$, *View*)

For example, a paper posted on a conference mini-track should be visible to track and conference chairs, but not necessarily to other track or mini-track chairs. Ancestors can be notified of new offspring, as an owner can be notified of new ancestors.

*Offspring role*. An entity created in a parent space was by definition created by an actor with the right to enter that space. If a space bans the owner of an object in it, the object is disowned, contradicting P1. A child object's owner must enter its space to act on it, even if they can't do anything else. By extension, they can also enter any ancestor space. This doesn't imply any other rights. The ACS principle is:

P13. An entity owner should be able to enter any ancestor space.

e.g. adding a mini-track paper should let one enter the track and conference spaces. Any space should allow its offspring owners to enter it:

$Role_{Offspring}$ = (**Offspring** , *Space* , *Enter*)

Table 24.13 summarizes the basic access rights for entities and spaces.

| **Entity** | *View* | *Delete* | *Edit* | *Display* | *Allocate* |
|---|---|---|---|---|---|
| *Ancestor* | √ | | | | |
| *Parent* | √ | | | √ [1] | |
| *Owner* | √ | √ | √ | √ [2] | √ |
| *LocalPublic* | √ [1,2] | | | | |
| | | | | | |
| **Space also** | *Enter* | *Create* | | | |
| *Ancestor* | √ | | | | |
| *Owner* | √ | √ | | | |
| *LocalPublic* | √[1] | √[1] | | | |

TABLE 24.13: Entity and space access rights. [1] As allocated by the owner. [2] As allocated by the parent.

## 24.4.12 The act of display

To display an object is to let others view it. The right to display isn't the right to view, e.g. viewing a video online doesn't let you display it on your web site[89]. Display is the meta-right to view, i.e. the right to give the right to view an object to others, e.g. privacy is the meta right to display the persona object. As people have private numbers in a phone book, so Facebook or Linkedin persona are displayed

89. A more complex example is: if going out in public implicitly gives others the right to view you, anyone can take a photo of you without your consent, but they can't display that photo on a magazine cover without your consent.

to the public by owner consent. The phone company that owns a phone book list can also choose not to display a listing, giving the ACS principle:

P14. Displaying an entity in a space requires both persona and space owner consent.

Displaying an item in a space is its owner giving display rights to the space owner. For example, to put a physical notice on shopkeeper notice board involves these steps:

1. *Creation*. Create a notice. You own it and can still change it, or even rip it up.
2. *Permission*. Ask the board owner if it can be posted on the notice board.
3. *Post*. The board owner either vets notices in advance or lets people post themselves.
4. *Removal*. As the notice is displayed by mutual consent, either can remove it. A poster can also ask that it be removed..

The shopkeeper's right to take a notice down isn't the right to destroy it, because *he or she doesn't own it*. Nor can he or she alter (deface) notices on the board.

The same social logic applies online. Create a video on YouTube gives you view rights to it, but it isn't yet displayed to the public, as this right belongs to the space owner. Giving the right to display YouTube video is like giving a notice to a shopkeeper to post on their board. The item owner gives the space owner the right to display it in their space. In general, to display any video, photo or text in any online space requires mutual consent, as one party gives another the right to display, giving the ACS principle:

P15. An entity owner must give view meta-rights to a space owner to display in that space

| **Display result** | | **Space owner** | |
|---|---|---|---|
| | | *Accept* | *Reject* |
| **Object owner** | *Submit* | **YES** | NO |
| | *Withdraw* | NO | NO |

TABLE 24.14: An open display interaction.

Display as a rights transaction is the basis of all publishing. Table 24.14 shows how the mutual interaction between authors and publishers, or object owners and space owners, operates. A space can delegate display rights, to let creators display as desired, e.g. YouTube. Or it may vet items before display and reject some, e.g. ArXiv, which also lets authors withdraw submissions. Bulletin boards let anyone submit but not withdraw, and reserve the right to moderate postings, i.e. reject later. Authors who publish must give all rights the publisher. An author can't "un-publish" a paper, but then again, neither can the publisher[90]. Usually the right to publish a work is given once only, but some publishers contract the right to do so many times, e.g. publishing one IGI book chapter led to its re-publication in other collections without author permission[91] (Whitworth and Liu, 2008).

*Entity creation*. Technically, creating an entity is simple - the program just creates it - but socially adding into another's space isn't a one-step act. Adding a YouTube video involves:

1. *Registration*. Create a YouTube persona.
2. *Entry*. Enter YouTube (not banned).

90. A journal can 'retract' a publication, to deny it, but can't 'un-publish' it.

91. Namely, in 'Selected Readings on the Human Side of Information Technology' and 'Human Computer Interaction: Concepts, Methodologies, Tools, and Applications'

3. *Creation*. Create and upload a video.

4. *Edit*. Edit video title, notes and properties.

5. *Submit*. Request YouTube to display the video to their public.

6. *Display*. The public sees it and can vote or comment.

YouTube lets anyone registered in the public role (1) enter their space (2) and create a video, by uploading or recording, which they own (3). They can view it in private and edit details (4). At this point, the video is visible to them and administrators, but not to the public. They can still delete it. It is then submitted to YouTube for display to its public (5). This occurs quickly as display rights are delegated (6). To create, edit and display a video are distinct steps. YouTube can still *reject* videos that fail its copyright or decency rules. This isn't a delete, as the owner can still view, edit and resubmit it. In contrast, a technology based design that lets space owners delete videos at will discourages participation, because people could waste their effort.

*Consistency*. For the above logic to be consistent, it should also apply when the video itself a space for dependent comments and votes. Indeed it is, as video owners have to choice to allow comments or votes just as YouTube had to right to accept their video (Figure 19). That YouTube gives the same rights to others as it takes for itself is a key part of its success and a basic principle of socio-technical designs.

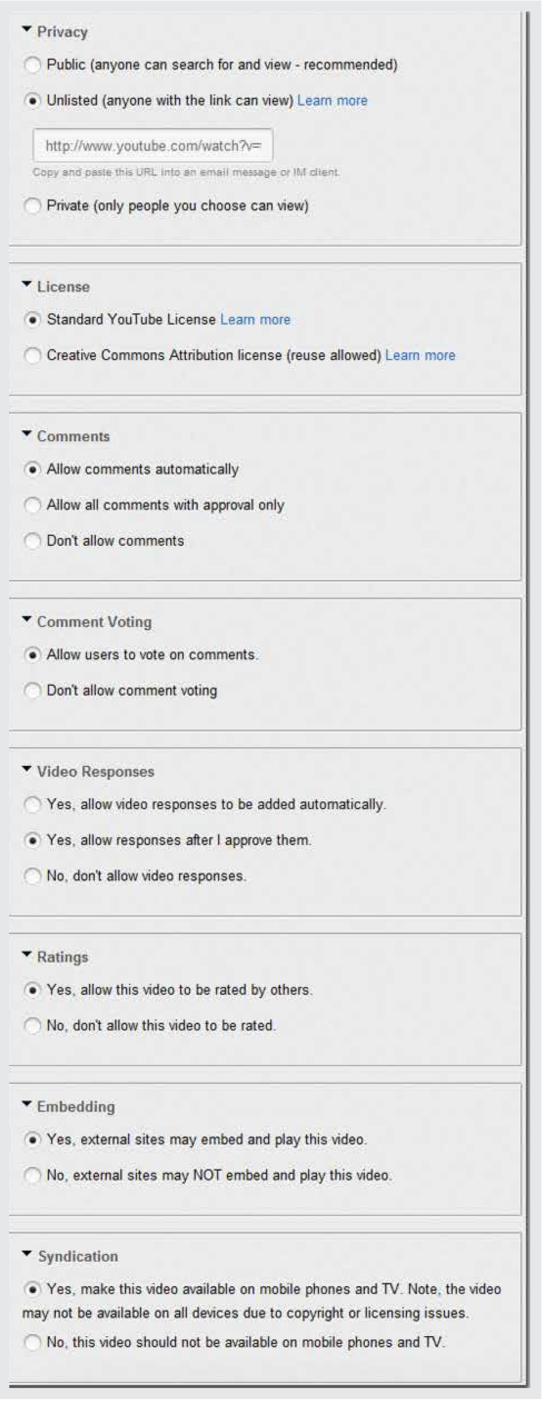



FIGURE 24.19: YouTube video rights.

### 24.4.13 Re-allocating rights

The right to re-allocate rights is part of social interaction. It allows socio-technical systems to evolve from an initial state of one administrator with all rights, to a community sharing rights. Use and meta rights can be re-allocated, as follows:

1. *Transfer*. Re-allocate all rights, including meta-rights. Rights are irrevocably given to the new owner, e.g. after selling a house, the old owner has no rights to it.

2. *Delegate*. Re-allocate use rights but not meta-rights. It can be reversed, e.g. renting.

3. *Divide*. A right divided among an actor set requires all to agree to permit an act, and any party can stop it, e.g. couples who jointly own a house.

*Multiply*. A right multiplied across an actor set lets them all exercise it as if they owned it exclusively, e.g. couples who severally share a bank account.

| | *Allocated by* | | *Allocated to* | |
|---|---|---|---|---|
| | Meta- rights | Use rights | Meta- rights | Use rights |
| *Transfer* | | | √ | √ |
| *Delegate* | √ | | | √ |
| *Divide use* | √ | ½ √ | | ½ √ |
| *Divide all* | ½ √ | ½ √ | ½ √ | ½ √ |
| *Multiply use* | √ | √ | | √ |
| *Multiply all* | √ | √ | √ | √ |

Table 24.15: Results use and meta rights re-allocations.

Dividing a right means that all must agree to it, while multiplying one means that any party alone can activate it. This isn't just splitting hairs, as if a couple owns house jointly, both must sign the sale deed to sell it, but if they own it severally, either party can sell it and take all the money. Re-allocating rights applies to many social situations, e.g. submitting a paper online can transfer all rights to a primary author, or also let them delegate rights to others, or divide rights so all authors must confirm changes, or multiply rights to all authors. Table 24.15 shows the resultant states. Each has different consequences, e.g. multiplying the edit right is risky but invites participation, while dividing it is safe but reduces contributions.

*Delegation*. Delegation, by definition, doesn't give meta-rights, so a delegatee can't pass rights on. Renting an apartment gives no right to sub-let, and lending a book doesn't give the right to on-lend it. It isn't hard to show that if delatees delegate, accountability is diluted. If one loans a book to one who loans it to another who loses it, who is accountable? This gives the operational principle:

P16. Delegating doesn't give the right to delegate.

Allocating use rights to an existing object makes the target person accountable for it, so it requires consent, e.g. one can't add a paper co-author without agreement. The principle is:

P17. Allocating existing object use rights to a person requires their consent.

An ACS might ask: "*Bob offers you edit rights to 'The 2012 Company Plan', do you accept?*" In contrast, rights to null acts, like view or enter, or to acts like create, can be allocated without consent because they imply no accountability:

P18. Allocating null rights to existing objects, or the right to create, requires no consent.

So space owners can freely delegate entry, view and create rights to anyone.

*Social networks*. Social networks currently send messages like:

*"X wants to be friends with you"*

In this tit-for-tat social trade: X offers to make you a friend if you make them one, i.e. it is a social trade. Yet by P7, one can befriend another without their permission.[92] If the software allowed it, we might get messages like:

*"X considers you a friend "*

This is giving friendship, not trading it. As one can love a child unconditionally, even if they don't return the favor, so friendship needn't be a commercial transaction.

For a social network to consider the friends of my friends also my friends contradicts P16. As liking someone doesn't guarantee that one will like their friends, so making a friend shouldn't reset my friend list. This illustrates a technical option that failed because it had no social basis.

### 24.4.14 Implementation

Traditional access control enforcement is done by a security kernel mechanism. A security kernel is a trusted software module that intercepts every access request call submitted to a system and decides if it should be granted or denied, based on some specified access policy model. Usually, a centralized approach is used, so one policy decision point handles all resource requests. The user sees either an executed action result or a permission denied message. SNSs have millions of users so centralized or semi-decentralized certificates are a bottle neck. This plus the social need for local ownership by content contributors suggests a strategy of distributed certificates to implement the ACS policy model outlined here. Allowing local policy decision points to handle resource requests also ensures local user control over resources. If distributed certificates are stored in the stakeholder's namespace, only he or she can access and modify them (Figure 24.20).

92. So giving another the right to *view* your wall doesn't let them spam you with change notices from theirs.

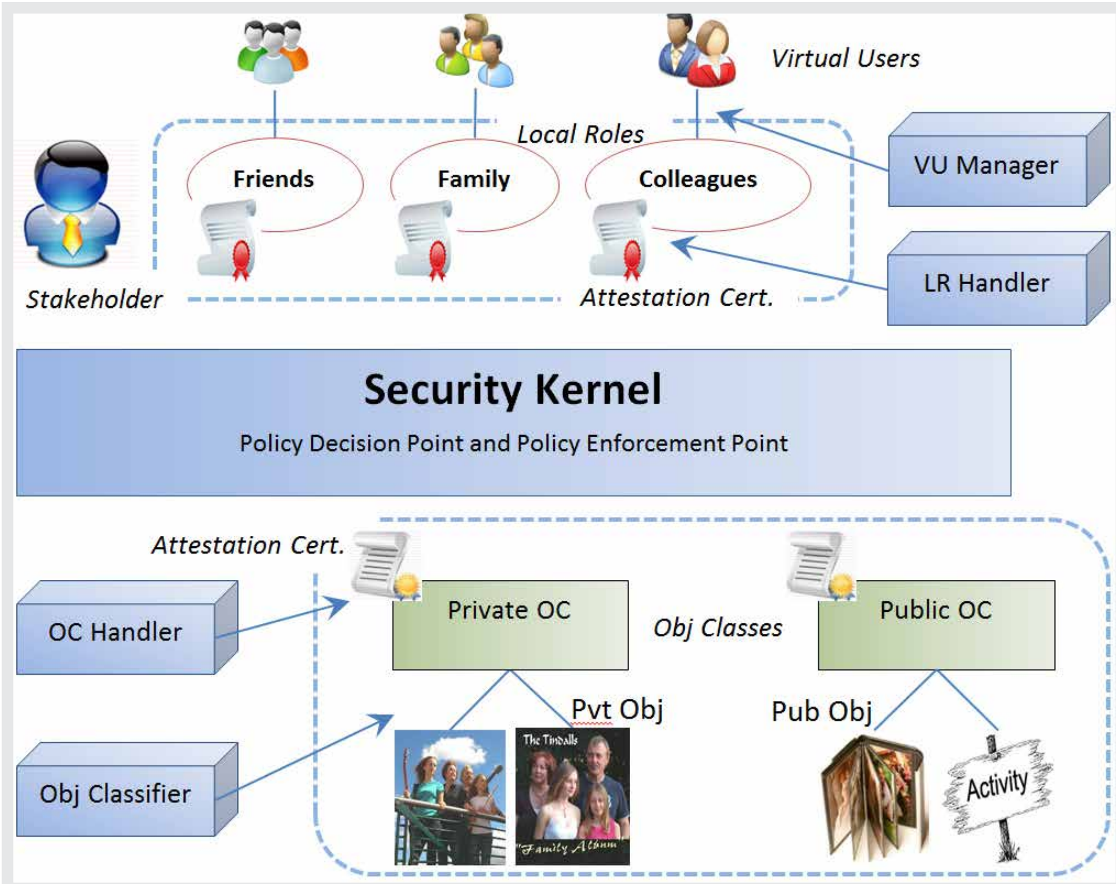



Figure 24.20: Distributed access control model architecture.



## 24.4.15 Summary

A legitimate ACS model can manage rights by assigning owner, parent, ancestor, offspring, and local public roles to objects and spaces. The ACS axioms deduced are:

1. *All non-null entity rights should be allocated to actors.*
2. *A persona should be owned by itself.*
3. *Every entity has a parent space, up to the system space.*

4. *Any right to use an object implies a right to view it.*
5. *Any communication act should have prior mutual consent.*
6. *A role is a right expressed in general terms, as a pointer or set.*
7. *A space owner can ban or give entry to a persona without its owner's permission.*
8. *A meta-right is the right to allocate any entity right, including itself.*
9. *Creation is always an act on a space, up to the system space.*
10. *The creator of new entity should immediately gain all rights to it.*
11. *A person can view in advance any rights that could apply to them.*
12. *A space owner should have the right to view any offspring.*
13. *An entity owner should be able to enter any ancestor space.*
14. *Displaying an entity in a space requires both persona and space owner consent.*
15. *To display an entity in a space, the entity owner gives view meta-rights to the space owner.*
16. *Delegating doesn't give the right to delegate.*
17. *Allocating existing object use rights to a person requires their consent.*
18. *Allocating null rights to existing objects, or the right to create, requires no consent.*

The above are social requirements not technical necessities, aiming at social sustainability. We are in the process of formalizing this model as a social interaction standard for any socio-technical system.

## 24.4.16 Discussion questions

Research selected questions from the list below. If you are reading this chapter as part of a class - either at university or a commercial course - you can research these questions in pairs and report back to the class, with reasons and examples.

1. What is access control? What types of computer systems use it? What don't? How does it traditionally work? How do social networks challenge this? How has access control responded?

2. What is a right in human terms? Is it a directive? How are rights represented as information? Give examples. What is a transmitted right called? Give examples.

3. What is the difference between a user and an actor? Contrast user goals and actor goals. Why are actors necessary for online community evolution?

4. Is a person always a citizen? How do communities hold citizens to account? If a car runs over a dog, is the car accountable? Why then is the driver accountable? If online software cheats a user, is the software accountable? If not, who is? Give an example. If automated bidding programs crash the stock market and millions lose their jobs, who is accountable? Can we blame technology for this?

5. Contrast an entity and an operation. What is a social entity? Is an online persona a person? How is a personae activated? Is this "possessing" an online body? Is a persona "really" you? If a program activates a persona, is it an online zombie? What online programs greet you by name? Do you like that? If an online banking web site welcomes you by name each time, does it build up a relationship? Who are you relating to?

6. Estimate how many hours a day you interact with technology. Be honest. Of those, how many are with online programs vs. people? Which do you prefer? Are any online programs your friend? Try out mobile phone help you can converse with, like Siri. Ask it to be your *personal* friend and report the conversation. If AI improved, would you like a personal AI friend?

7. Must all rights be allocated? What rights must be? Why? What manages online rights? Are AI programs accountable for rights allocated to them? In the USS Vincennes tragedy, was the computer program that shot down the Iranian civilian airliner held to account? Why not? What caused the error? What changed afterwards?

8. Who should own a persona and why? For three STSs, create a new persona, use it to connect, try to edit it, then to delete it. Compare what properties you can and can't change. If you delete it entirely, what remains? Can you resurrect it? Describe two ways to join an online community. Which is easier? More secure?

9. Describe, with examples, current technical responses to the social problems of persona abandonment, transfer, delegation and orphaning. What do you recommend in each case?

10. Why is choice over displaying oneself to others important for social beings? What is the right to control this called? Who has the right to display your name in a telephone listing? Who has the right to remove it? Does the same apply to an online registry listing? Investigate three online cases and report what they do.

11. How do information entities differ from objects? How do spaces differ from items? What is the object hierarchy and how does it arise? What is the first space? What operations apply to spaces but not items? What

operations apply to items but not spaces? Can an item become a space? Can a space become an item? Give examples.

12. How do comments differ from messages? Define the right to comment as an AEO triad. If a comment becomes a space, what is it called? Demonstrate with three commenting STSs. For systems that allow "deep" commenting (comments on comments on comments, etc), what is going on? (Look at who adds). Would a chat type conversation function be simpler than so many indents?

13. For each operation set below, explain the differences, give examples, and give another variant:

    - *Delete*: Delete, undelete, destroy.
    - *Edit*: Edit, append, version, revert.
    - *Create:* Create.

What is the difference between create and edit? Define a fourth operation set.

14. Is viewing an object an act upon it? Is viewing a person an act upon them? How is viewing a social act? Can viewing an online objects be a social act? Why is viewing necessary for social accountability?

15. What is communication? Is an information transfer a communication, e.g. a download? Why should communication require mutual consent? What happens if it isn't mutual? How does opening a channel differ from sending a message? Can a sender be anonymous to a receiver? Can a receiver be anonymous to a sender? Can senders or receivers be anonymous to the transmission system? Describe online systems that enable channel control.

16. Answer the following for a landline phone, mobile phone and Skype: How does the communication request manifest? What information does a receiver get and what choices do they have? What happens to anonymous senders? How does one create an address list? What else is different?

17. What is a role? Can it be empty or null? How is a role like a maths variable or computing pointer? Give role examples from three popular STSs. For each, give the ACS triad, stating what values vary. What other values could vary? Use this to suggest new useful roles.

18. How can roles, by definition, vary? For three different STSs, describe how each role variation type might work. Give three different examples of implemented roles and suggest three future developments.

19. If you unfriend a person, should they be informed? Test and report what actually happens on three common SNs. Must a banned bulletin board "flamer" be notified? What about someone kicked out of a chat room? What is the general principle here?

20. What is a meta-right? Give physical and online examples. How does it differ from other rights? Is it still a right? Can an ACS act on meta-rights? Are there ACS meta-meta-rights? If not, why not? What then does it mean to "own" an entity?

21. Why can't an ACS creating an item be an act on that item? Why can't it be an act on nothing? What then is it an act upon? Illustrate with online examples.

22. Who owns a newly created information entity? By what social principle? Must this always be so? Find online cases where you create a thing online but *don't* fully own it.

23. In a space, who, initially, has the right to create in it? How then can others create in that space? What are creation conditions? What is the jus-

tification? Illustrate object, operation, access, visibility and edit conditions. How does transparency apply?

24. Give three examples of creating an entity in a space. For each, specify the owner, parent, ancestors, offspring and local public. Which role(s) can the owner change?

25. For five different STS genres, give examples of online creation conditions. Create something in each. Was the result transparent? Find two examples of non-transparent creations.

26. For the following, explain why or why not. Suppose you are the chair of a computer conference with several tracks. Should a track chair be able to exclude you, or hide a paper from your view? Should you be able to delete a paper from their track? What about their seeing papers in other tracks? Should a track chair be able to move a paper submitted to their track by error to another track? Investigate and report comments you find on online systems that manage academic conferences.

27. An online community has put an issue to a member vote. Evaluate these STS options:

    a. Voters can see how others voted, by name, before they vote.

    b. Voters can see the vote average before they vote.

    c. Voters can only see the vote average after they vote, but before all voting is over.

    d. Voters can only see the vote average after all the voting is over.

Find online votes to illustrate. Do the same for these voting options:

Voters aren't registered, so one person can vote many times.

Voters are registered, but can change their one vote any time.

Voters are registered, and can only vote once, with no edits.

Can the person calling the vote legitimately define these vote conditions? What if they set conditions like all votes must be signed and will be made public?

1. Is posting a video online like posting a notice in a local shop window? Explain, covering permission to post, to display, to withdraw and to delete. Can a post be deleted? Can it be rejected? Explain the difference. Give online examples.

2. Give physical and online examples of rights re-allocations. Specify rights and meta-rights. If four authors publish a paper online, list the ownership options. Discuss how each might work out in practice. Which would you prefer and why?

3. Should delegating give the right to delegate? Explain, with physical and online examples. What happens to ownership and accountability if delegatees can delegate? Discuss a worst case scenario.

4. If a property is left to you in a will, can you refuse to own it, or is it automatically yours? What rights can't be allocated without consent? What can? Which of these rights can be freely allocated: Paper author. Paper co-author. Track chair. Being friended. Being banned. Bulletin board member. Logon ID. Bulletin board moderator. Online Christmas card access? Which require receiver consent?

5. Investigate how SN connections multiply. For you and four friends, list the number of friends and the average. Based on this, estimate the total possible friends of friends in general. By looking at your friend's friend lists, give, in your case, the friends of friends actual. Estimate how many messages or notifications you get from all your friends per week. From that, estimate the average messages per friend per day. So if you friended

all your friend's friends, potentially, how many messages could you expect per day? What if you friended your friend's friend's friends too? Why is the number so large? Discuss the film, Six Degrees of Separation.

6. Demonstrate how to "unfriend" a person in three social networks. Are they notified? Is unfriending "breaking up"? That an "anti-friend" is an enemy, suggests "anti-Facebook" sites. Investigate technology support for people you hate, e.g. celebrities or my relationship ex. Try anti-organization sites, like sickfacebook.com. What purpose could technology support for anti-friendship serve?

## 24.5 PART 5: THE FUTURE

**"**The future isn't technical or social but both**"**

### 24.5.1 Technology utopianism

Technology utopianism is the belief that technology alone creates the future. It is popular in fiction, e.g. Rosie in The Jetsons, C-3PO in Star Wars and Data in Star Trek are robots that read, talk, walk, converse, think and feel. As we do these things easily, how hard could it be? In films, robots learn (Short Circuit), reproduce (Stargate's replicators), think (The Hitchhiker's Guide's Marvin), become self-aware (I, Robot) and eventually replace us (The Terminator, The Matrix). In this view, computers are an unstoppable evolutionary juggernaut (Figure 24.21), but right now they couldn't conquer a planet of cockroaches.

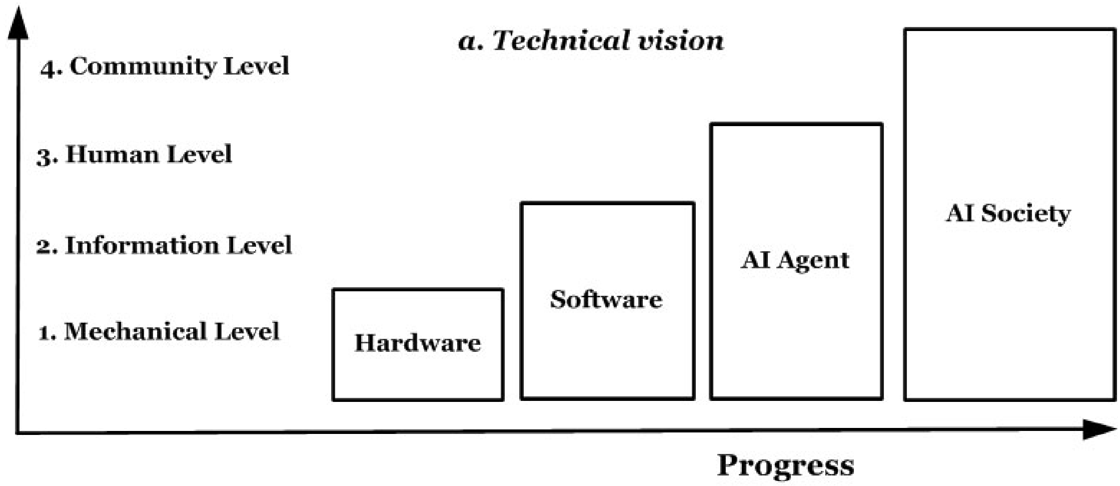



Figure 24.21: Technological utopianism.



Despite cartoons and science fiction, today's housework robot is the Roomba (Figure 24.22). It picks up dirt, but user problems include turning in circles, paralysis, thread and hair clogs and getting stuck. Even washing it in water kills it. Make no mistake, this is a clever robot, yet it needs help to survive *in a house with flat terrain, no enemies and free food.*[93] Dropped outside or left alone, it can't survive. Even insects do better.

93. The Roomba 'feeds' from electricity sockets in the house.

Courtesy of Fleischer Brothers production. Copyright: pd (Public Domain (information that is common property and contains no original authorship)).

*Courtesy of Brayon Michael Pieske. Copyright: pd (Public Domain (information that is common property and contains no original authorship)).*




Figure 24.22 A-B-C: Robot help in: a. Superman, b. Star Wars, c. Reality (the first iRobot Roomba version).

Conversation is the same. People chat easily about many things, but the computers next to us are strangely silent. The technology exists already so why don't cars talk to us like KITT in Knight Rider? Is it because computer conversations are embarrassing?[94]

94. Try a computer conversation at http://cleverbot.com/. Here is a sample, run through twice:
*If you become me, am I you?*
I believe in you babe, or, What if I don't believe they are?
*If you are not you, who are you?*
I am not a vampire, or, I'm you. *How do you know?*
Because I'm a girl, or, Because you know I know I know you know me.
*But girls can be vampires.*
Yes they can, or, No they can't.

Spatial coordination is the same. Compare the Robot World Cup and the human World Cup. Artificial Intelligence (AI) advocates say robots will surpass people in forty years, but they said that forty years ago. Like a mirage, AI breakthroughs are always forty years hence — or is it forty thousand years? For vision, computers still struggle with retinal let alone cortical activity:

*"Computers are no real competition for the human brain in areas such as vision, hearing, pattern recognition and learning. … And when it comes to operational efficiency there is no contest at all. A typical room-size supercomputer weights roughly 1,000 times more, occupies 10,000 times more space and consumes a millionfold more power …"*

*-- Boahen, 2005*

The point isn't what computers can't do, but that tasks like talking, walking and thinking aren't as easy as they might seem. Technology utopianism predict a "singularity" based on Moore's law, that computer processing power

*So could you be a vampire girl?*
No. I am not, or, A girl.
*Could you become a vampire girl?*
Why? or, No, could you become a girl?
*If you were bitten by a vampire would you be a vampire girl?*
Who have you told that I am Jesus? or, I'd buy an X-wing, because they're awesome.
*So you could become a vampire girl?*
I doubt it. But I know it could be true, or, I could be a lot of things.
*That contradicts what you said before.*
I don't think it does, or, You are contradicting yourself.
*You could be not you and a vampire.*
No, I'm just a rabbit, or, Girl vampire.
*Could you be not you as a vampire girl?*
No, or, I could probably be a giraffe, if science were able to make me so.
*You just said you could be a vampire girl which is not you.*
I never said I was a boy, or I never said I was a girl. This painful conversation has no end.

doubles every eighteen months[95]. It says that shortly, super-intelligent computers will replace people (Kurzweil, 1999). This “big lie”[96] fantasy, sees the future as just more of the same processing computers already have. Yet evolution is never more of the same, and the brain isn’t just a big computer.

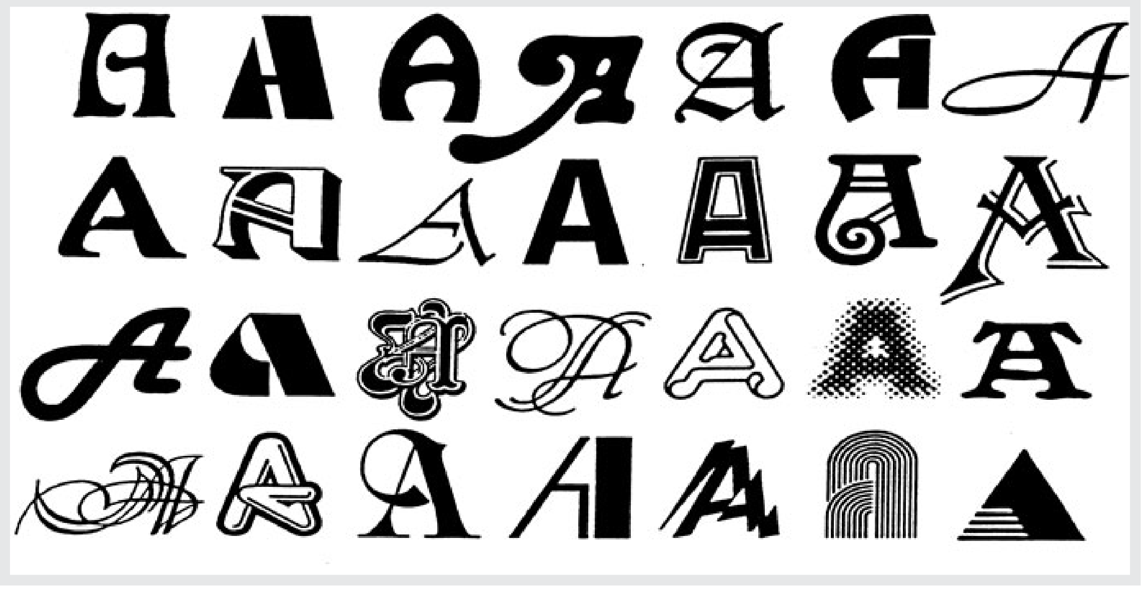


Figure 24.23: Letraset page for letter ‘A’

*Copyright status: Unknown (pending investigation). See section “Exceptions” in the copyright terms.*

Computers calculate better than us, as cars travel faster and cranes lift more, but calculating isn’t all the brain does. Simple processing[97] works for simple cases, but real tasks like vision, hearing, thinking and conversing are *productive*, i.e. their information increases geometrically with size[98]. The productivity of language is

95. See for example http://karlnordstrom.ca/ideas/?p=6

96. A ‘big lie’ is a statement so ludicrous it is assumed to be true, e.g. the statement that all people are equal, when diversity is an obvious principle of natural selection. Of course, all should have equal rights.

97. Simple processing works only at the informational level, e.g. a literal number or word recall, number calculations or a ‘photographic’ memory of a scene.

98. Even on an 8x8 chess board, the number of possible chess games is $10^{120}$ - more than the atoms in the universe. In an Indian tale, the inventor of chess was offered a boon by the king. He asked for a grain of

that five year olds can speak more sentences than they could learn in a lifetime at a sentence per second (Chomsky, 2006). Children easily see that a Letraset page (Figure 24.23) is all 'A's, but computers struggle with such productive variation. Using pixel level processing for pattern recognition is: "*like trying to understand bird flight by studying only feathers. It just cannot be done.*" (Norman, 1990). AI experts who saw beyond the hype knew decades ago that productive tasks like language wouldn't be solved anytime soon (Copeland, 1993).

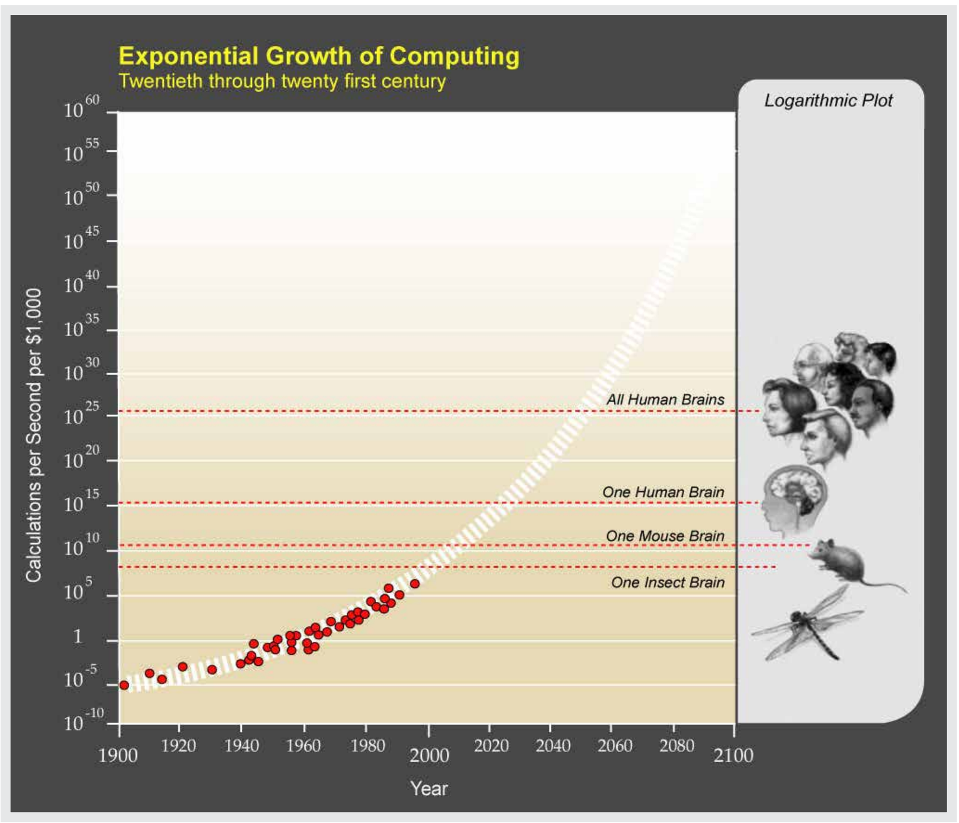



Figure 24.24: The exponential growth of simple process power.



wheat on the first chessboard square, two grains on the second, four on the third, eight on the fourth, and so on, doubling the grains each time. It seemed a modest request, so the king agreed, but the result was over 18 billion, billion grains, more weight than all life on Earth. This is the productivity problem.

The bottom line for simple processing is *the 99% barrier*, e.g. 99% accurate computer voice recognition makes one error per 100 words, but an error per minute is well below conversation standards. For computer auto-drive cars, 99% accuracy is an accident a day! In the 2005 DARPA Grand Challenge, five of 23 autonomous vehicles finished a simple course (Miller et al, 2006). In 2007, six of eleven better funded vehicles finished an urban track with a top average speed of 14mph. Yet skilled people drive for decades on harder roads, in worse weathers, in heavier traffic, and faster, with *no* accidents[99]. The brain didn't cross the 99% performance barrier just by increasing simple processing power.



99. A good MTBA (Mean Time Between Accidents) is twenty years, see http://ridingsafely.com/ridingsafely3.html



Figure 24.25 A-B: Leftmost: Kim Peek was the inspiration for the film, Rain Man. Rightmost: Dustin Hoffman in the role of Rain Man.

How can a brain handle "incalculable" tasks? It *is* an information processor. Its trillion (1012) neurons are biological on/off devices powered by electricity that allow logic gates (McCulloch and Pitts, 1943), i.e. in principle no different from transistors. If processing power really depends on neuron/transistor numbers, computers should be at the brain's potential soon. Figure 24.24 suggests that computers processed as an insect in 2000, as a mouse in 2010, will be as a human in 2020 and beyond all humans in 2045. Of course this is nonsense, as right now computers can't even do what ants do, with a neuron sliver. Or bees or cockroaches or flying beetles. How will they then jump to conversation, pattern recognition and learning in a few decades?

The reason is that calculating power wasn't the answer to incalculable tasks, as our brain, in its evolution, discovered. In *savant syndrome*, people who can calculate 20 digit prime numbers in their head need full time care to live in society, e.g. Kim Peek, who inspired the movie Rain Man, could recall every word on every page of over 9,000 books, including all Shakespeare and the Bible, but had to be cared for by his father (Figure 24.25). He was neurologically disabled, as later parts of his brain didn't arise.

Savants then are the brain working without its more recent sub-systems. That they calculate *better* suggests that the brain tried simple processing power and evolved past it. In contrast, technology utopians still don't see that more of the same isn't evolution.

Computers are *electronic savants*, calculation wizards that need minders to survive in the real world. If computers excel at the sort of processing the brain outgrew a million years ago, how are they the future? If super-computers built from PC video cards running in parallel are the future of computing, then bigger oxes are the future of farming! How can AI surpass HI (Human Intelligence) if it isn't even going in the same direction?

A system's performance isn't just its parts but also how they connect. Computers today follow von Neumann's architecture, but the brain didn't, e.g. it has no CPU (Sperry and Gazzaniga, 1967). It crossed the 99% performance barrier by taking design risks von Neumann avoided (Whitworth, 2009c). The processing of processing is avoided by computer science as it gives infinite loops, yet it allows symbolism - linking one brain neural assembly (a symbol) to another (a perception). This is the basis of meaning and language. Processing changes information so assumes a context[100]. Only by the processing of processing can we modify contexts, i.e. learn. Denying computers this option denied them meaning.

100. The context of information is the option set chosen from.

Rather than an inferior biological version of today's computers, the brain is a *different kind of processor altogether*. It processes its own processing to give language, mathematics and philosophy. The answer to the productivity problem wasn't more processing but the processing of processing. By this risky step, the brain perceives a "self", "others", "friends" and "community", the same constructs that human and computer savants struggle with. If today's super-computers aren't even in the same processing league as the brain[101], technology utopians are computing traditionalists posing as futurists.

### 24.5.2 The socio-technical vision

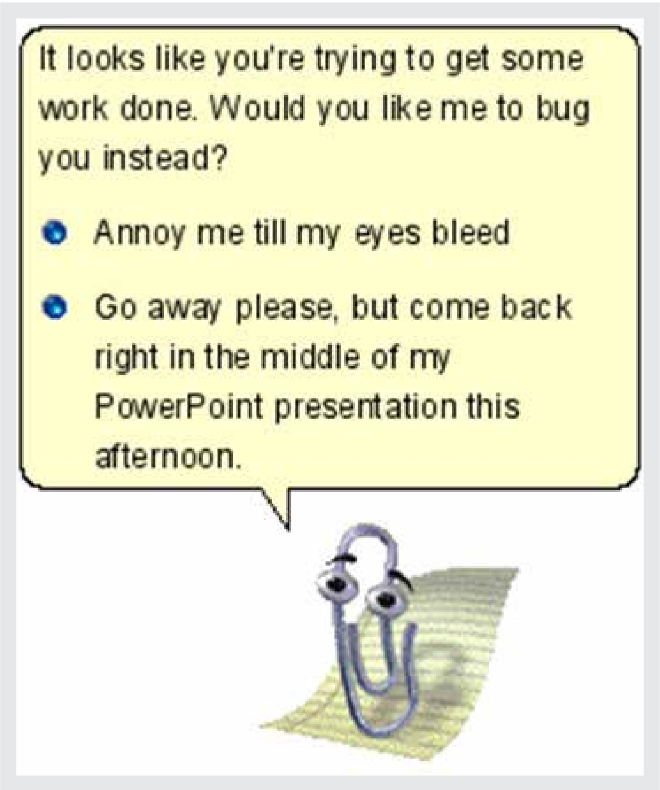



Figure 24.26: Mr. Clippy takes charge.



101. This is not to say they cannot be, just that to do this would require a basic change in their architecture.

The question facing computing isn't when it will replace people but when people will see it for what it is, e.g. Mr. Clippy, Office 97's paper clip assistant (Figure 24.26):

> "*It wouldn't go away when you wanted it to. It interrupted rudely and broke your train of thought.*"
>
> -- *Pratley, 2004*

Responses included "*Die, Clippy, Die*!" (Gauze, 2006), but its Microsoft designer still wondered: "*If you think the Assistant idea was bad, why exactly*?" The specific answer is: because it thought it was in charge. In Windows XP, Mr. Clippy was replaced by tags smart enough to know their place. Software that tries to be "smart" by itself quickly ends up like the sorcerer's apprentice.

Why tie up twenty-million-dollar super-computers to try to do what brains already do, with millions of years of real life beta-testing? Even if we redesign computers to work like the brain, say as neural nets, who is to say they won't inherit the same weaknesses? If the brain has solved the productivity problem as well as can be expected, lets change the goals of computing, from human mimicry to human assistance.

This is already happening. Driverless cars are still a dream but reactive cruise control, range sensing and assisted parallel parking already exist (Miller et al, 2006). Computer surgery struggles but computer supported remote surgery and computer assisted surgery are here today. Robots run clumsily but people with robotic limbs are more than able. Computer piloted drones are a liability but remotely piloted drones are an asset. Computer generated animations are great, but state-of-the-art animations like Avatar combine human actors and computers. Chess players advised by computers perform better than either alone[102]. In killer applications of the last decade, from email to Facebook, people do what they do best and technology do what it does best, e.g. email transmits information and people create meaning. So “horses for courses” is letting computers process information and people process meaning. That meaning is a level above information, implies that people should “mind” computers and computers shouldn’t control people.

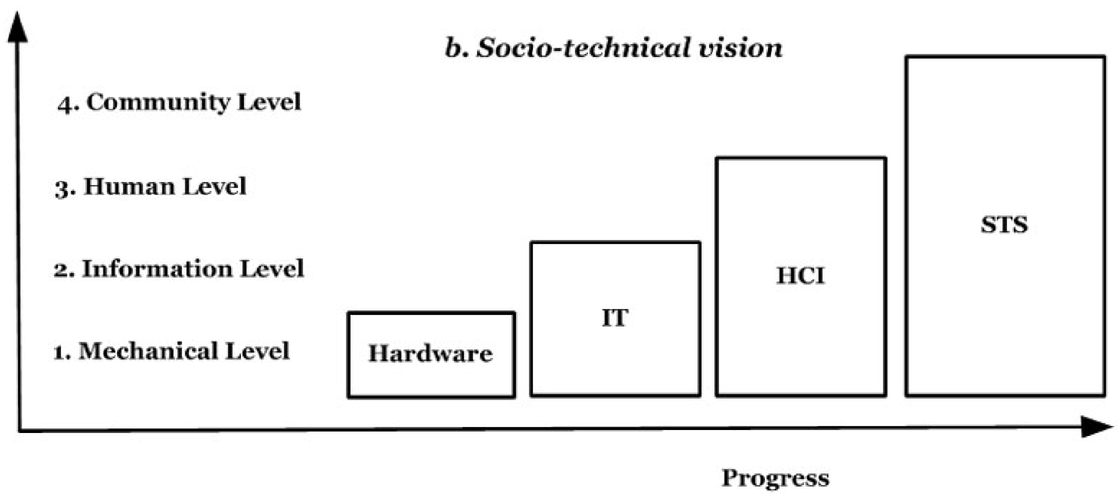



Figure 24.27: The socio-technical vision.

*Copyright status: Unknown (pending investigation). See section “Exceptions” in the copyright terms.*

Socio-technology is about technology *and* people, with the latter the “elder” system (Figure 24.27). If people direct technology it *may* go wrong, but if technology directs people it *will* go wrong. Higher levels directing lower ones is evolution, but lower ones directing higher ones is devolution. To focus on lower levels, because they are easier, isn’t progress[103]. To see the Internet only in technical terms

102. See Kasparov’s ‘The Chess Master and the Computer’, 2010

103. A man was looking for his lost keys at night under a well lit lamp post. When asked where he lost them, he replied: ‘Over there in the bushes - but the light is better here.’ Seeking intelligence in information is the same.

is to underestimate it, again! Let computers be background not foreground, as pervasive and ubiquitous computing theories propose. Technology should merge with people, not the other way around. Technology without a human context isn't even useless - it is pointless. If "technology is the future", something mindless and heartless is in charge of us. So the future is socio-technology not technology.

Some say the Internet is making us stupid[104] but a mirror just reflects. Online media showing human brutality, corruption or stupidity just reveal what is. The Internet, as a microscope and telescope on humanity, *is showing us to us*. It isn't physical, but thoughts cause words and deeds as guns fire bullets. Humanity's thoughts are now online for us to choose. We, the human race, are choosing what we think and what we think is now online, with web-counters keeping the score. What the Internet electronic mirror shows isn't always pretty but it is real and to change oneself one must first see oneself. The evolution of computing is a part of human evolution, of a social experiment that has been ongoing for thousands of years. Only by personal evolution, by seeing beyond ourselves, do we help it succeed.

### 24.5.3 Discussion questions

Research selected questions from the list below. If you are reading this chapter as part of a class - either at university or a commercial course - you can research these questions in pairs and report back to the class, with reasons and examples.

1. What is technology utopianism? Give examples from movies. What is the technology singularity? In this view, why must computers take over from people? What is the false assumption here?

2. What technology advances did the last century expect by the year 2000? Which ones are we are still awaiting? What do people expect robots to be doing by 2050? What is realistic? How do robot achievements like the

104. For example: 'The internet is full of idiots writing rubbish for other idiots to read.'; 'The internet is full of idiots and one of them might just be you.'; and 'Do not feed the trolls (DNFTT)

Sony dog rank? How might socio-technical design improve the Sony dog? In the socio-technical paradigm, how will robots evolve? Give examples.

3. If super-computers achieve the processing power of one human brain, then many brains, are many people together more intelligent than one? Review the "Madness of Crowds" theory, that people are less intelligent together. Give examples. Why doesn't adding more programmers to a project always finish it quicker? What, in general, affects whether parts perform better together? Is a super computer, with as many transistors as the brain has neurons, its processing equal? Explain.

4. How do today's super computers increase processing power? List the processor cores of the top ten? Which use NVidia PC graphic board cores? How is this power utilized in real computing tasks? How does processing cores operating in sequence or parallel affect performance? How is that decided in practice? (CS students only).

5. Review the current state-of-the-art for automated vehicles, whether car, plane, train, etc. Are any fully "pilotless" vehicles currently in use? What about remotely piloted vehicles? When does full computer control work? When doesn't it? (hint: consider active help systems). When might full computer control of a vehicle be useful? Suggest how computer control of vehicles will evolve, with examples.

6. What is the 99% barrier? Why is the last 1% of accuracy a problem for productive tasks? Give examples from language, logic, art, music, poetry, driving and another. How common are such tasks in the world? How does the brain handle them?

7. What is a human savant? Give examples past and present. What tasks do savants do easily? Can they compete with modern computers?

What tasks do savants find hard? What is the difference? Why do savants need support? If computers are like savants, what support do they need?

8. Find three examples of software that, like Mr. Clippy, thinks it knows best. Give examples of: 1. Acts without asking, 2. Nags, 3. Changes secretly, 4. Makes you work.

9. Think of a personal conflict you would like advice on. Keep it simple and clear. Now try these three options. In each case explain and ask the question the same way:

   a. Go to your bedroom alone, put a photo of family member you like on a pillow. Explain and ask the question out loud, then imagine their response.

   b. Go to an online computer like http://cleverbot.com/ and do the same.

   c. Ring an anonymous help line and do the same.

Compare and contrast the results. Which was the most helpful?

1. A rational way to decide is to list all the options, assess each one and pick the best. How many options are there for these contests: 1. Checkers, 2. Chess, 3. Civilization (a strategy game), 4. A MMORPG, 5. A debate. Which ones are computers good at? What do people do if they can't calculate all the options? Can a program do this? How do online gamers rate human and AI opponents? Why? Will this always be so?

2. Mr. Clippy was based on Bayesian logic. What data drove his decisions? What was left out? Why did users find him rude? Why couldn't he recognize rejection? What users liked Mr. Clippy? Turn on the auto-correct in Word and try writing the equation: i = 1. Why does Word get it wrong?

How can you fix it without turning off auto-correct? Give online examples of recommending and taking charge.

3. What is the difference between syntax and semantics in language? What are programs good at? Look at text-to-speech systems, like here, or translators here. How successful are they? Are computers doing what people do? At what level is the translating occurring? Are they semantic level transformations? Discuss John Searle's Chinese room thought experiment.

## 24.6 ACKNOWLEDGEMENTS

Thanks to the first author's wife for helpful advice, to the students of 158729 (STS Design) at Massey University for trying out the questions. Also thanks to Yijing Qian for Figure 24.2 and Figure 24.9.

## 24.8 REFERENCES

Ackerman, Mark S. (2000): *The Intellectual Challenge of CSCW: The Gap Between Social Requirements and Technical Feasibility. In Human-Computer Interaction, 15 (2) pp. 181-203*

Ahmad, Adnan and Whitworth, Brian (2011): *Distributed access control for social networks. In 2011 7th International Conference on Information Assurance and Security IAS, pp. 68-73*

Alberts, Bruce, Bray, Dennis, Lewis, Julian, Raff, Martin, Roberts, Keith and Watson, James D. (1994): *Molecular Biology of the Cell 3E. Garland Science*

Alexander, Christopher (1964): *Notes on the Synthesis of Form. Harvard University Press*

Alter, Steven (1999): *A general, yet useful theory of information systems. In Communications of the AIS, 1 (3)*

Beer, David and Burrows, Roger (2007): *Sociology and, of and in Web 2.0: Some Initial Considerations. In Sociological Research Online, 12 (5) pp. 1-15*

Benkler, Yochai (2002): *Coase's Penguin, or, Linux and "The Nature of the Firm. In Yale Law Journal, 112 (3) pp. 369-446*

*Berners-Lee, Tim (2000): Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web. HarperBusiness*

*Bertalanffy, Ludwig Von (1968): General System Theory: Foundations, Development, Applications (Revised Edition). George Braziller Inc*

*Boahen, Kwabena (2005): Neuromorphic Microchips. In Scientific American, 292 (5) pp. 56-63*

*Borenstein, Nathaniel S. and Thyberg, Chris A. (1991): Power, Ease of Use and Cooperative Work in a Practical Multimedia Message System. In International Journal of Man-Machine Studies, 34 (2) pp. 229-259*

*Boutin, Paul (2004): Can e-mail be saved. In Infoworld, 14*

*Burk, Dan L. (2001): Copyrightable functions and patentable speech. In Communications of the ACM, 44 (2) pp. 69-75*

*Callahan, David (2004): The Cheating Culture: Why More Americans Are Doing Wrong to Get Ahead. Mariner Books*

*Campbell-Kelly, Martin (2008): Historical reflections: Will the future of software be open source?. In Communications of the ACM, 51 (10) pp. 21-23*

*Chomsky, Noam (2006): Language and Mind. Cambridge University Press*

*Chung, Lawrence, Nixon, Brian A., Yu, Eric and Mylopoulos, John (1999): Non-Functional Requirements in Software Engineering (THE KLUWER INTERNATIONAL SERIES IN SOFTWARE ENGINEERING Volume 5). Springer*

*Clark, D. D. and Wilson, D. R. (1987): A Comparison of Commercial and Military Computer Security Policies. In: IEEE Symposium on Security and Privacy 1987 1987. pp. 184-195*

*Cohen, Bram (2003): Incentives Build Robustness in BitTorrent. In Workshop on Economics of PeertoPeer systems, 6 (22)*

*Copeland, Jack (1993): Artificial Intelligence: A Philosophical Introduction. Wiley-Blackwell*

*Cysneiros, L. M. and Leite, Julio Cesar Sampaio do Prado (2002): Non-functional requirements:from Elicitation to modeling languages. In Computer, 35 (3) pp. 8-9*

*David, Julie Smith, McCarthy, William E. and Sommer, Brian S. (2003): Agility: the key to survival of the fittest in the software market. In Communications of the ACM, 46 (5) pp. 65-69*

*Davis, Fred D. (1989): Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. In MIS Quarterly, 13 (3) pp. 319-340*

*Department of Defense (1985). TCSEC - Trusted Computer Security Evaluation Criteria (TCSEC), DOD 5200.28-STD. Retrieved 19 May 2012 from Department of Defense:*

*Diamond, Jared M. (1998): Guns, Germs, and Steel: The Fates of Human Societies. W. W. Norton and Company*

*Esfeld, Michael (1998): Holism and analytic philosophy. In Mind, 107 (426) pp. 365-380*

*Ferraiolo, David F. and Kuhn, D. Richard (2004): Role Based Access Control. In Review Literature And Arts Of The Americas, 14 (5) pp. 554-563*

*Figart, Deborah M. and Golden, Lonnie (eds.) (2000): Working Time: International Trends, Theory and Policy Perspectives (Routledge Advances in Social Economics). Routledge*

*Forman, Bruce Jay and Whitworth, Brian (2007): Information Disclosure and the Online Customer Relationship. In: Quality, Values and Choice Workshop, Computer Human Interaction 2007, Portland, Oregon, USA. pp. 1-7*

*Freeden, Michael (1991): Rights (Concepts in Social Thought).*

*Freudenthal, Eric, Pesin, Tracy, Port, Lawrence, Keenan, Edward and Karamcheti, Vijay (2002): dRBAC: distributed role-based access control for dynamic coalition environments. In Proceedings 22nd International Conference on Distributed Computing Systems, pp. 411-420*

*G., Geen, R. and Gange, J. J. (1983): Social facilitation: Drive theory and beyond. In: Blumberg, Herbert H. (ed.). "Small Groups and Social Interaction: v. 2 (Small Groups & Social Interactions)". John Wiley and Sons Ltdp. 141–153*

*Gediga, Gunther, Hamborg, Kai-Christoph and Duntsch, Ivo (1999): The IsoMetrics Usability Inventory: An Operationalization Of ISO 9241-10 supporting summative and formative evaluation of software systems. In Behaviour and Information Technology, 18 (3) pp. 151-164*

*Hoffman, L. R. and Maier, N. R. F. (1961): Quality and acceptance of problem solutions by members of homogenous and heterogenous groups. In Journal of Abnormal and Social Psychology, 62 pp. 401-407*

*Johnson, Deborah G. (2001): Computer Ethics (3rd Edition). Prentice Hall*

*Jonsson, Erland (1998): An integrated Framework for Security and Dependability. In Information Security, pp. 22-29*

*Kant, Immanuel (1999): Critique of Pure Reason (The Cambridge Edition of the Works of Immanuel Kant). Cambridge University Press*

*Karp, Alan H., Haury, Harry and Davis, Michael H. (2009): From ABAC to ZBAC : The Evolution of Access Control Models From ABAC to ZBAC : The Evolution of Access Control*

*Models. In Control, (0) pp. 22-30*

Keeney, Ralph L. and Raiffa, Howard (1976): *Decisions with Multiple Objectives: Preferences and Value Tradeoffs. Cambridge University Press*

Kelly, Erin (ed.) (2001): *Justice as Fairness: A Restatement. Belknap Press of Harvard University Press*

Kienzle, Darrell M. and Wulf, William A. (1998): *A practical approach to security assessment. In: Proceedings of the 1997 workshop on New security paradigms 1998. pp. 5-16*

Knoll, Kathleen and Jarvenpaa, Sirkka L. (1994): *Information technology alignment or in highly turbulent environments: the concept of flexibility. In: Proceedings of the 1994 computer personnel research conference on Reinventing IS managing information technology in changing organizations managing information technology in changing organizations 1994. pp. 1-14*

Kurzweil, Ray (1999): *The Age of Spiritual Machines: When Computers Exceed Human Intelligence. Viking Adult*

Lampson, B. W. (1969): *Dynamic protection structures. In: Proceedings of the November 18-20, 1969, fall joint computer conference 1969. pp. 27-38*

Lessig, Lawrence (1999): *Code and Other Laws of Cyberspace. Basic Books*

Lindquist, Christopher (2005): *Fixing the requirements mess. In CIO, 0*

Locke, John (1963): *An essay concerning the true original extent and end of civil government: Second of "Two Treatises on Government" (1690). In:* Somerville, John and Santoni, *Ronald (eds.). "Social and Political Philosophy: Readings From Plato to Gandhi". Anchorp. 169–204*

Lorenz, E. N. (1963): *Deterministic Nonperiodic Flow. In Journal of the Atmospheric Sciences, 20 (2) pp. 130-141*

Losavio, Francisca, Chirinos, Ledis, Matteo, Alfredo, Levy, Nicole and Ramdane-Cherif, *Amar (2004): Designing Quality Architecture: Incorporating ISO Standards into the Unified Process. In IS Management, 21 (1) pp. 27-44*

Mandelbaum, Michael (2002): *The Ideas that Conquered the World: Peace, Democracy, and Free Markets in the Twenty-first Century. PublicAffairs*

McCulloch, Warren S. and Pitts, Walter H. (1943): *A logical calculus of the ideas immanent in nervous activity. In Bulletin of Mathematical Biophysics, 5 (4) pp. 115-133*

*MessageLabs (2006). The year spam raised its game; 2007 predictions. Retrieved 19 May 2012 from MessageLabs:*

*MessageLabs (2010). Intelligence Annual Security Report, 2010. Retrieved 19 May 2012 from MessageLabs:*

*Meyrowitz, Joshua (1985): No Sense of Place: The Impact of Electronic Media on Social Behavior. Oxford University Press, USA*

*Miller, George A. (1956): The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information. In Psychological Review, 63 pp. 81-97*

*Miller, Isaac, Garcia, Ephrahim and Campbell, Mark (2006): To Drive Is Human. In IEEE Computer, 39 (12) pp. 52-56*

*Mitchell, William J. (1995): City of Bits: Space, Place, and the Infobahn (On Architecture). The MIT Press*

*Moreira, Ana, Araújo, João and Brito, Isabel (2002): Crosscutting quality attributes for requirements engineering. In Proceedings of the 14th international conference on Software engineering and knowledge engineering SEKE 02, (0)*

*Norman, Donald A. (1990): The Design of Everyday Things. New York, Doubleday*

*Nuseibeh, Bashar and Easterbrook, Steve (2000): Requirements engineering: a roadmap. In: Proceedings of the Conference on The Future of Software Engineering 2000. pp. 35-46*

*OECD (1996). Guidelines for the Security of Information Systems. Retrieved 19 May 2012 from OECD:*

*Penrose, Roger (2005): The Road to Reality : A Complete Guide to the Laws of the Universe. Knopf*

*Pinto, Jeffrey K. (2002): Project Management 2002. In Research Technology Management, 45 (2) p. 22–37*

*Porra, Jaana and Hirschheim, Rudy (2007): A Lifetime of Theory and Action on the Ethical Use of Computers: A Dialogue with Enid Mumford. In Journal of the Association for Information Systems, 8 (9) pp. 467-478*

*Poundstone, William (1992): Prisoner's Dilemma. Anchor*

*Raymond, Eric S. (1999): The Cathedral & the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary. OReilly Media*

*Regan, Priscilla M. (1995): Legislating Privacy: Technology, Social Values, and Public Policy. University of North Carolina Press*

*Reid, Fraser J. M., Malinek, Vlastimil, Stott, Clifford J. T. and Evans, Jonathan ST. B. T. (1996): The messaging threshold in computer-mediated communication. In Ergonomics, 39 (8) pp. 1017-1037*

*Ridley, Matt (2010): The Rational Optimist: How Prosperity Evolves. Harper*

*Rosa, Nelson S., Justo, George R. R. and Cunha, Paulo R. F. (2001): A framework for building non-functional software architectures. In Parallel Computing, pp. 141-147*

*Rose, E. (2000): Balancing internet marketing needs with consumer concerns: a property rights framework. In ACM SIGCAS Computers and Society, 30 (2) pp. 20-24*

*Samuelson, Pamela (2003): Unsolicited communications as trespass?. In Communications of the ACM, 46 (10) pp. 15-20*

*Sanai, Hakim Abu L Majd Madud (1968): The Enclosed Garden of Truth. Theophania Publishing*

*Sanders, Mark S. and McCormick, Ernest J. (1993): Human Factors In Engineering and Design. McGraw-Hill Science*

*Seabold, Daniel E., Honemann, Daniel H. and Balch, Thomas J. (eds.) (1993): Robert's Rules of Order Newly Revised, 11th edition. Da Capo Press*

*Shannon, Claude E. and Weaver, Warren (1949): The Mathematical Theory of Communication. University of Illinois Press*

*Shannon, Claude E. and Weaver, Warren (1971): The Mathematical Theory of Communication. University of Illinois Press*

*Shirky, Clay (2008): Here Comes Everybody: The Power of Organizing Without Organizations. Penguin Press*

*Short, John, Williams, Ederyn and Christie, Bruce (1976): Visual communication and social interaction - The role of 'medium' in the communication process. In The Social Psychology of Telecommunications, pp. 43-60*

*Simone, Mauricio De and Kazman, Rick (1995): Software architectural analysis: an experience report. In CASCON 95 Proceedings of the 1995 conference of the Centre for Advanced Studies on Collaborative research,* Skinner, Burrhus F. (1948): 'Superstition' in the pigeon. In *Journal of Experimental Psychology*, 38 (2) pp. 168-172

Smith, Heather A., Kulatilaka, Nalin and Venkatramen, N. (2002): Developments in IS practice III: Riding the wave: extracting value from mobile technology. In *Communications of the Association for Information Systems*, 8 (0) pp. 467-481

Sommerville, Ian (2004): Software Engineering (9th Edition). Addison Wesley

Spence, Robert and Apperley, Mark (2011). Bifocal Display. Retrieved 6 June 2013 from [URL to be defined - in press]

Sperry, R. W. and Gazzaniga, M. S. (1967): Language following surgical disconnexion of the hemispheres. In: Millikan, Darley (ed.). "Brain Mechanism Underlying Speech and Language". Grune and Stratton

Tenner, Edward (1997): Why Things Bite Back: Technology and the Revenge of Unintended Consequences (Vintage). Vintage

Thompson, Mary, Johnston, William, Mudumbai, Srilekha, Hoo, Gary, Jackson, Keith and Essiari, Abdelilah (1999): Certificate-based Access Control for Widely Distributed Resources. In *Proceedings of 8th USENIX Security Symposium*, pp. 215-228

Toffler, Alvin (1980): The Third Wave. Bantam

Weiss, Aaron (2003): Ending spam's free ride. In *netWorker*, 7 (2) pp. 18-24

Whitworth, Brian (2009b): The social requirements of technical systems. In: Whitworth, Brian and Moor, Aldo de (eds.). "Handbook of Research on Socio-Technical Design and Social Networking Systems (2-Volumes)". Information Science Reference

Whitworth, Brian (2011): The Virtual Reality Conjecture. In *Prespacetime Journal*, 2 (9) p. 1404–1433

Whitworth, Brian (2009a): A Comparison of Human and Computer Information Processing. In: Pagani, Margherita (ed.). "Encyclopedia of Multimedia Technology and Networking (2 Volume Set)". Idea Group Publishingp. 230–239

Whitworth, Brian (2006): Measuring disagreement. In: Reynolds, Rodney A., Woods, Robert and Baker, Jason D. (eds.). "Handbook of Research on Electronic Surveys and Measurements".

Whitworth, Brian and Bieber, Michael (2002): Legitimate Navigation Links. In: *ACM Hypertext 2002, Demonstrations and Posters* 2002, Maryland, USA. pp. 26-27

*Whitworth, Brian and Friedman, Robert S. (2009): Reinventing Academic Publishing Online. Part I: Rigor, Relevance and Practice. In First Monday, 14 (8)*

*Whitworth, Brian and Liu, Tong (2008): Politeness as a Social Computing Requirement. In: Luppicini, Rocci (ed.). "Handbook of Conversation Design for Instructional Applications (Premier Reference Source)". Information Science Referencepp. 419-436*

*Whitworth, Brian and Liu, Tong (2009): Channel E-mail: A Sociotechnical Response to Spam. In IEEE Computer, 42 (7) pp. 63-72*

*Whitworth, Brian and Moor, Aldo de (2002): Legitimate by Design: Towards Trusted Virtual Community Environments. In: HICSS 2002 2002. p. 213*

*Whitworth, Brian and Whitworth, Alex P. (2010): The social environment model: Small heroes and the evolution of human society. In First Monday, 15 (11)*

*Whitworth, Brian and Whitworth, Elizabeth (2004): Spam and the Social-Technical Gap. In Computer, 37 (10) pp. 38-45*

*Whitworth, Brian, Bañuls, Victor, Sylla, Cheickna and Mahinda, Edward (2008): Expanding the Criteria for Evaluating Socio-Technical Software. In IEEE Transactions on Systems, Man, and Cybernetics, 38 (4) pp. 777-790*

*Whitworth, Brian, Gallupe, Brent and McQueen, Robert (2000): A cognitive three-process model of computer-mediated group interaction. In Group Decision and Negotiation, 9 (5) pp. 431-456*

*Whitworth, Brian, Gallupe, Brent and McQueen, Robert (2001): Generating agreement in computer-mediated groups. In Small Group Research, 32 (5) pp. 625-665*

*Wright, Robert (2001): Nonzero: The Logic of Human Destiny. Vintage*

## About the authors

### Brian Whitworth

Personal Homepage:
http://brianwhitworth.com/

Born in England and brought up in New Zealand, Brian Whitworth currently works at Massey University in Auckland, New Zealand. After doing a mathematics degree, and a Master's thesis on split-brain neuropsychology, Brian joined the New Zealand Army, where he was the first specialist to complete regular army officer cadet training. He worked as an army psychologist, and then in computer operational simulations (wargames), while simultaneously raising four wonderful children, until he retired in 1989 as a Major. Brian then completed his doctorate on online groups, and students at his university used the social voting system he built until the World Wide Web arrived. In 1999, he worked in the USA as a professor, and published in journals like Small Group Research, Group Decision and

Negotiation, Communications of the AIS, IEEE Computer, Behavior and Information Technology, and Communications of the ACM. More recently, he was the senior editor of the Handbook of Research on Socio-Technical Design and Social Networking Systems, written by over a hundred leading experts worldwide. His interests include computing, psychology, quantum theory and motor-cycle riding.

### Adnan Ahmad

Has also published under the name of:
"Adnan Ahmad", "Adnan Ahmed", and "A. Ahmad"

Adnan Ahmad was born in Lahore, Pakistan. He received his Bachelor of Science (Hons.) from Govt. College University (GCU, 2005), with a major in software engineering, and a Masters of Science from Lahore University of Management Sciences (LUMS, 2009), with a major in distributed systems. He worked for five years in industry, getting hands on experience with cutting edge software and hardware technologies. His PhD in Information Technology at Massey University, New Zealand, was on a formal model of distributed rights allocation in online social interaction. He has published in well-known conferences like Worldcomp, IFIP SEC, IAS, IAIT and Trustcom. His current research applies socio-technical design principles to computer security, and his other interests include crowd sourcing, distributed systems, spam and image processing.

## YOUR NOTES AND THOUGHTS ON CHAPTER 24

Record your notes and thoughts on this chapter. If you want to share these thoughts with others online, go to the bottom of the page at:
http://www.interaction-design.org/encyclopedia/socio-technical_system_design.html

**NOTES**

# Chapter 25

# 25. Semiotics

## and Human-Computer Interaction

## *by Clarisse Sieckenius de Souza.*

This chapter covers why and how Semiotics can help advance some of the major goals of Human-Computer Interaction (HCI) and be useful when designing interactive products. It begins with brief definitions and explanations of a few central concepts in Semiotics. This is followed by a discussion of harder challenges involved in bringing Semiotics into the domain of HCI research and the consequences of viewing *computers as media*. Following *Semiotic Engineering* concepts, which we have been developing and using for two decades now, we then revisit computer-mediated communication in view of 21$^{st}$ century literacy issues. First, we show that basic computing skills exhibited by contemporary users are in fact *semiotic engineering* abilities of the same sort as required from professional designers. Then we show how these skills can leverage an individual's participation in a variety of social processes. In conclusion, the chapter presents our personal answer to the question that most readers certainly have in mind: '*So, what's in it for me?*'

## 25.1 INTRODUCTION

Semiotics and HCI have more in common than is usually acknowledged in either side of the cultural divide that has been keeping them apart. Researchers and professional practitioners working with these disciplines have different interests and perspectives when selecting, a0pplying and building *knowledge*. As a result, mutual understanding has not only been rare, but usually perceived as "more cost than benefit".

This chapter is an essay on why and how Semiotics can help advance some of the major goals in HCI. It begins with a definition of Semiotics and a brief explanation of a few central concepts that will be used throughout the chapter. Next, it discusses some deeper challenges of using Semiotics in the context of HCI research. Then it explores the notion of *computers as media*, the hallmark of all semiotic approaches to HCI proposed to date. It highlights the interdisciplinary work of pioneers like Mihai Nadin and Peter Bøgh Andersen and ends with a description of *Semiotic Engineering*, a comprehensive semiotic theory of HCI which we have been developing and using at SERG, the Semiotic Engineering Research Group in Rio de Janeiro, since 1990. In subsequent sections, we take a closer look at computer-mediated communication in view of contemporary opinions that having basic programming skills is as important for citizens of the 21$^{st}$ century as reading, writing and counting have been in the 20$^{th}$ century and before. Some examples of basic computing literacy skills exhibited by contemporary users are framed as *semiotic engineering* abilities of the same sort as required from professional HCI designers. This helps us to show why computers *are* indeed the most pervasive *media* used by contemporary societies, and it points at computing literacy issues with which HCI is not only necessarily involved, but also in which it has clear vested interests. The chapter then provides our answer to the provocative question implied by its title, proposed by the editors of the *Encyclopedia of HCI*: "So, what's in it for me?" The answer, I hope, will attract more attention to this fascinating discipline in times when computers so obviously coalesce and transform society's means of communication and participation.

## 25.2 SOME PRELIMINARY DEFINITIONS

*Semiotics* is the study of signs. Although strictly correct, this definition is not helpful for those who do not know what signs are and how they can be studied. So, let us begin with additional definitions. The selected material will also give the reader a flavor of the *cultural divide* between Semiotics and HCI as expressed by terminology and conceptual framing.

In the *Encyclopedia of Semiotics* (Bouissac 1998), the entry for "sign" explores numerous variations in the way philosophers and semioticians have defined the central object of interest in this discipline over the centuries. Here are two of them:

a. In the tradition inaugurated by Ferdinand de Saussure: *"[...] the sign can be understood as a* **correlation of differences**.*"* (Bouissac 1998: p. 573)

b. In the tradition inaugurated by Charles Sanders Peirce: *"[...] the defining characteristic of signs is their capacity* **to determine additional signs** *[in the mind]."* (Bouissac 1998: p. 574)

Saussure and Peirce are the *founding fathers* of Semiotics. Although they were contemporaries (Saussure died in 1913 and Peirce in 1914), they lived in different continents (Europe and North America, respectively) and followed completely independent paths. Saussure was a linguist, interested in a formal characterization of natural languages. Peirce was a logician, interested in meaning and knowledge discovery processes. The definitions above are both very true to their respective theories, but they are also mysterious for non-semioticians. I picked them up because they also illustrate rhetorical differences between disciplines, which we have to understand in order to reap the benefits of interdisciplinary research.

### 25.2.1 Saussure

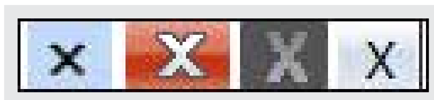


Figure 25.1: Oppositions in *form* that do not correspond to oppositions in *meaning*.

*Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.*

Keeping in mind that Saussure (Saussure 1972) was focused almost exclusively on linguistic signs, the *correlates* with which he defined the sign were: an acoustic image (the *'signifiant'*, or signifier); and a concept (the *'signifié'*, or signified). The former is a physical entity, whereas the latter is an abstract one. *Differences* determine what constitutes a sign at various levels and dimensions. Just as a brief illustration, take the word 'privacy' in English. Its acoustic image in spoken language is not unique: some say |'prɪvəsi|, while others say |'praɪvɪsi|. Although in English the long vowel |aɪ| is different from the short vowel |ɪ| (*e. g.* |baɪt| and |bɪt| are not the same word), this difference is neutralized in words like 'privacy': The concept, no matter the pronunciation, is the same. So, what is a sign? In phonology, |aɪ| and |ɪ| are signs because they have significant *differences*. One is associated with the concept of 'long vowel', and the other with 'short vowel'. However, in *syntax* or *semantics* |'prɪvɪsi| and |'praɪvɪsi| are not distinct signs.

Difference, as the reader can infer, is a very powerful principle in meaning making and sense making. In natural languages, differences and oppositions are *systemic*. That is, certain opposing combinations are recurrent and meaningful in the language, whereas others aren't. Therefore, a language can be described as an inventory of signs and sign combinations (or structures), for which the *correlation* between the *signifiant* and the *signifié* can be formally established by systemic differences. Here is a simple example of how this principle can be applied in the context of HCI.

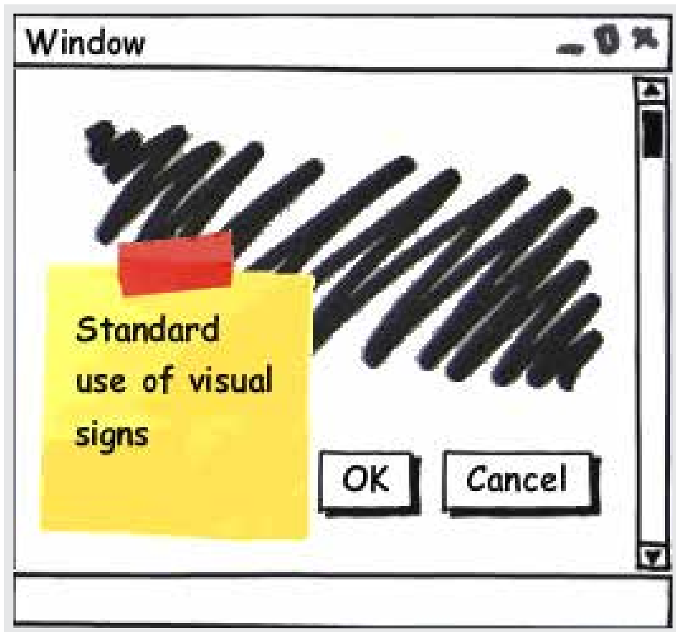



Figure 25.2: Well-formed sign structures in HCI.



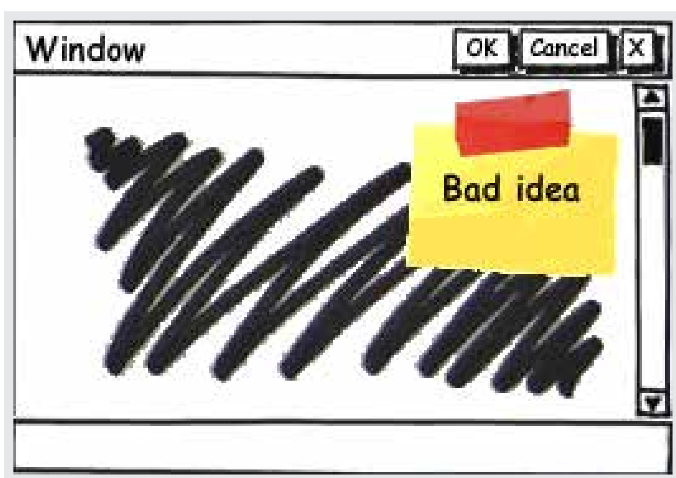



Figure 25.3: Ill-formed sign structures in HCI.

In Figure 25.1, we see different renditions of a familiar window control element. The *signifiants* constitute different signs in the equivalent of *HCI phonology*. Nevertheless, in HCI *syntax* and *semantics*, the whole collection corresponds to a single lexical item (a single sign for 'close window'). In Figure 25.2 and Figure 25.3, we see the parallel with Saussure's patterns of structures. In one case, the 'close window' sign is combined with other signs that are systemically used in conjunction with it, whereas in the other case it is combined with signs that should never co-occur with it. Therefore, in this interface language, only one of the two structures is well-formed.

Note that *signifiant/signifié* correlations in interface languages are much weaker than in natural languages. Semiotic correlations are *programmed* into software. Compared to natural language, this can be a blessing and a curse. The blessing is that in interface languages, unlike in natural languages, the conceptual correlate of signs can (and must) be fully specified and instantiated by a mechanical process. Thus, we always have access to the origin of meaning. The curse is that mechanical instantiation of the *signifiés* as correlates of interface *signifiants* is arbitrarily programmed by human minds. That is, nothing can prevent a mischievous programmer from specifying that '|x|' causes a window to freeze, rather than close. This is not the case in natural languages: neither you nor I can *decide and establish* that from now on 'privacy' means something else in English. The correlation between parts of signs in formal grammatical descriptions of natural languages is not established by individuals.

This example gives us a chance to mention another fundamental dichotomy in Saussurean theory, that between *langue* (language) and *parole* (speech). This significant opposition has been posed to account for the fact that individual variations in language *use* do not affect *the system*. It is easy to collect numerous examples of individual linguistic behavior *contradicting* the general rules and conventions of a given language system (*e. g.* we make occasional grammatical mistakes as we speak, which may be due to heavy cognitive loads). However, individual and sporadic variations do not change or affect the language *system*. It is only if and when such variations make their way into a formal abstract and general sign system, extending over large spans of space and time, that they actually cause a change in *langue*. Other variations, within smaller spatial and temporal coverage, correspond to changes in *parole*.

Saussure's interest was in *a theory of langue*, his formal theory of signs. This refers us to a silent Cartesian tradition running under the surface of non-Cartesian concepts, such as the determining forces of social structures and conventions. The Cartesian gene in the theory is the assumption that *linguists* can describe *langue* and that individual minds can have access to supra-individual concepts and formulate them *in abstract*, free of psychological, social, spatial and temporal contingencies. In other words, a linguist's description of *langue* objects should be warranted by this individual's ability to capture abstract '*signifiés*' that exist independently of his or her mind and to preserve them from contamination of individual interpretation when building formal theories of language.

Part of Descartes' much more complex philosophy, not discussed here, was that individual minds could have access to ultimate meanings by exercising methodical doubt and analysis (Descartes 2004). Eventually, Descartes believed, the mind would reach an unquestionable state, some primary cognition that constitutes the true meaning of the object(s) under consideration. In the remainder of this chapter, whenever we use the word *Cartesian* we are referring to this specific aspect of Descartes' theory, an important one for discussing alternative semiotic theories and their application to HCI.

### 25.2.2 Peirce

Peirce (Peirce 1992), as already mentioned, was interested in logic and the origin of meaning. His theory can actually be framed as a long and deeply elaborate rebuttal to the Cartesian canons we just mentioned (Santaella 2004). He defined signs as a three-part structure consisting of a representation (the *representamen*), an *object* and a mediating element (the *interpretant*) that binds the object to its representation in somebody's mind. In other words, unless there is a mental mediation, nothing is a *representation of* anything. This will probably remind the reader of the proverbial saying that "meaning is in the mind of the beholder". But, as the next paragraphs will show, the theory goes far beyond intuitive interpretations of the old adage.

The definitional feature of signs mentioned in (b) on page 2 — that they can generate other signs — refers to the fact that, according to Peirce, the mediating *interpretant*, just like the *representamen*, **stands for** the object that it binds to a particular representation. In other words, the *interpretant* is a second-order representation of the object itself. The sign has, in itself, a generative recursive seed that produces other signs, *ad infinitum*. For example, let us go back to the word 'privacy'. This is a sign where |'prɪvɪsi| is the *representamen* and something out-there-in-the-world is its *object* because in some minds (like yours and mine) there is an *interpretant* that binds them together. Is this *interpretant* unique? Is it the same for everyone? Is it the same for you or me throughout our entire lives? The answer is 'no'. From the very moment that we inspect this *interpretant* (*i. e.* that we take it to mean the *object* out-there-in-the-world which we call |'praɪvɪsi|), it becomes a second-order representation of its very *object*. A newly generated *interpretant* (*e. g.* an explanation or further elaboration referring to the *object*) is instantiated, which generates yet another sign instantiation as soon as our mind engages in further interpretation.

In order to illustrate how these concepts can be used in a practical HCI context description, let us go back to the '|x|' window control (see Figure 25.4). In a Peircean account of this sign, '|x|' is the *representamen* of the whole sign, whereas the specified computational procedure that causes a well-determined subsequent state of the system (in which the 'Street video' window is no longer visible) is its *object*. All of the interesting things have to do with the *interpretant*. What mediating sign (which will generate other signs) is binding representation and object together? To answer the question, we would have to inspect *minds*, like the programmers' and the users', for example.

An exhaustive account for *all interpretants* — all possible reasons and related facts that can cause '|x|' to *stand for* whatever it is that makes the window disappear when we click on it — in programmers' and users' minds, over large spans of time and space, is beyond the ability of individual minds, caught in the contingencies of momentary interpretation.

We can already see how Peirce's theory differs in focus and essence from Saussure's, and also why, in the context of this chapter, a semiotic theory's positioning relative to certain aspects of Cartesian tradition can extensively determine its fate in HCI. For instance, the tension between a Peircean perspective on meaning and the algorithmic nature of interface sign interpretation and generation processes in computer systems cannot be missed. Neither can the apparently smooth compatibility between Saussure's theory and the basic tenets of formal language and automata theory in Computer Science (Hopcroft and Ullman 1979) be ignored.

Nevertheless, there is more promise in Peircean theory for HCI than meets the eye. Let us briefly introduce a couple of additional concepts and definitions.


FIGURE 25.4: How simple interface signs can generate unlimited semiosis.



The first is Peirce's most widely known and used definition of sign: "A *sign* [...] is something which stands to somebody for something in some respect or capacity. It [...] creates in the mind of that person an equivalent sign, or perhaps a more developed sign." (Buchler 1955: p. 99) Note that the mediation of the *interpretant* is done 'in some respect or capacity' (*i. e.* for some potentially partial and arbitrary reason). Thus, although most users will tend to generate an interpretation for '|x|' that is at least superficially appropriate for the immediate context of interaction sketched in Figure 25.4, subsequently generated signs are likely to be very different. For example, thanks to some unfortunate previous experience, one user may accept that '|x|' means 'close the window', but develop further interpretations according to which the presence of '|x|' in that particular window is actually the result of negligent programming. She may believe that if she clicks on it and closes that window, the browser will freeze. In her mind, to prevent the unfortunate situation, the '|x|' control should not be active, especially because the 'Close & Return' button is there to control the closing operation safely. Some other user, however, thanks to luckier interactions, may develop the interpretation that the '|x|' control and the 'Close & Return' button *mean* exactly the same thing. The difference, he may think, is a *practical* one. The generic 'close a window' can be communicated with different signs: to click on '|x|' or press some specific key combination. The latter can be faster than the point-and-click alternative. Thus, whereas advanced users may *interpret* the presence of '|x|' as meaning that the use of keyboard input to control the window on Figure 25.4 is allowed, novice users may miss this signification of '|x|' altogether and interpret that to close the window they must click on the 'Close & Return' button. The next steps in each user's *theories* of what '|x|' means in different situations are virtually impossible to predict since they depend substantially on the users' experience with software, their level of technical awareness, and virtually all other meanings that populate their minds.

The second concept of Peircean theory that this example helps to illustrate is the process of *abductive reasoning* (also known as hypothetical reasoning or abduction), which underlies — according to the theory — all meaning-making activities. Very simply put, this kind of reasoning consists of generating hypotheses to explain (*i. e.* to interpret) significant elements of reality around us. Once a hypothesis is generated, it is tested against ready-to-hand evidence. If the evidence contradicts the hypothesis, a new one is generated and tested. If not, the hypothesis is accepted as a general principle capable of binding not only the tested instances of representations to the tested objects, but also — and more importantly — *future* instances and yet-to-be-met objects as well. The principle is true until further evidence is found that calls for *revision* of beliefs. This indefinitely long process of sign generation is named *semiosis*. It can, for example, explain Carroll and Rosson's observations reported in an influential early work in HCI, *The paradox of the active user* (Carroll and Rosson 1987). The authors remark that users "hastily assemble *ad hoc* theories" of how the system works and refuse to learn more even though they hold important misconceptions about the interface language semantics, which can lead to inefficient or ineffective interaction. A Peircean account of this phenomenon tells us that, in the process of abductive reasoning, users come upon mediating signs that do not require or motivate further abductive effort. They are content with their inferencing and fixate a generative belief (*i. e.* an *interpretant* that will be used in other sense-making situations). Even if from a programmer's point of view the users' reasoning and interpretations are totally *wrong*, for as long as they 'work', the users believe that they are *right*. The meaning of 'work', and not so much the meaning of interface elements themselves, is the key to understanding interaction patterns and choices in this case.

The third and last additional concept that we will discuss with respect to Peircean Semiotics is that of a/the *true meaning* of signs. If human minds can engage in widely (and wildly) different interpretive directions, how is it that two people can understand each other, or at least truly believe that they do so? And, much more importantly, is there such a thing as a/the *true meaning* of signs?

In a Peircean perspective, meaning emerges in a constantly changing chain of *"equivalent [...] or perhaps more developed" signs*. Subsequent *interpretants* are connected, corrected, adjusted, expanded, stabilized and destabilized by *the insistence of the real*. Without going into the very complex details of Peirce's theory of truth and true meanings, it is worth highlighting a few points about his particular view of reality and reality's role in sign-making processes:

I. Sign making (which for the purposes of this chapter is the same as *sense making*) is a natural disposition of the human species, just like flying is a natural disposition of birds. Humans cannot help but assign meaning to whatever they sense in the world around them. Therefore, reality is *always* mediated by signs.

II. Because it is an abductive inferencing process, human instinctive sign making is fundamentally prone to error. We constantly make mistakes while interpreting reality. However, our innate interpretive apparatus is such that it can *correct* interpretations in the presence of new evidence. This mode of mental operation constitutes the first and more essential step in *any* sort of human mental activity, from the mundane daily inferences about the freshness of products on market stalls to the scholarly debate about the validity of scientific theories.

III. The principle that makes human knowledge and understanding gravitate towards 'truth' is the *insistence of reality*, the inexorable massive set of evidence provided by the world around us. This is also an indication that, in this theory, truth is not the result of an individual's introspective activity (as Descartes would wish), but a collective ongoing semiotic process, in which abductive corrections made and expressed by other minds penetrate culture through social processes that eventually affect individual sense-making. Hence, as a social, cultural and historical process, human interpretation gravitates toward truth.

Back to HCI, these three points allow for a reinterpretation of heated controversies. Take for example Jared Spool's criticism of User-Centered Design. In a CHI Panel in 2005 (Spool and Schaffer 2005), among other things, Spool claimed that "beyond small teams with 'simple' issues, formalized UCD doesn't seem to work" and that "UCD pretends to act like an engineering discipline (formalized methods that have repeatable results independent of practitioner), but actually behaves as if it's a craft (dependent completely on skills and talents of practitioners with no repeatable results)." (p. 1174)

Spool's position is a good example of how the "insistence of reality" is related to, but not the same as, "repeatable results". A Peircean account of meaning tells us that predictions of how users will interpret interface signs must be taken very cautiously. When making sense about *the natural and the cultural world*, a user's abductive reasoning paths will eventually be corrected and refined by the *insistence* of a massive volume of facts that penetrate the mind as signs contingently related to each other "in some respect or capacity". Contingency — the *craft* that lies at the opposite of "repeatable results" expected by Spool — is actually the essence of human sense making.

Computer encoded *interface* signs, however, are different. They are, and must be, uniquely defined and *engineered* into software artifacts that continually "repeat" exactly the same mechanical interpretation and generation for each and every representation in the program. These representations are produced by human minds (the software designers' and developers') and *meant to affect* other human minds (the users'). The inexorable repetition of these representations in computer systems' interfaces is ontologically *very* different from what Peirce refers to as *the insistence of the real*. Actually, computer interface signs are no more than the expression of a particular moment in someone's (or some group of people's) abductive reasoning path, which is as prone to interpretive error as any user's abductions while trying to make sense of the interface.

We thus see that Spool's vigorous criticism of UCD echoes some of the Cartesian notions we discussed above. He, just like UCD adopters, is looking for repeatable semiosis "independent of practitioner", meaning independent of mind, which should nevertheless be mentally achieved by an individual's use of rigorous methods of inspection and analysis. The interest of a semiotic perspective on this debate is to show that there are theoretical choices beyond Cartesianism, and that depending on whether the focus is on human semiosis or computer semiosis, some choices are better than others. For illustration, we comment on two alternative paths.

Louis Hjelmslev (Hjelmslev 1961) followed Saussure's ideas, but gave them a more *axiomatic* interpretation. In his view, signs could be studied as the correlation between any two autonomous systems. Neither one had to (although they might) be anchored in psychological reality. Charles Morris (Morris 1971), in turn, followed Peirce's ideas, but gave them a radically behaviorist interpretation. In his view, signs should be defined in terms of stimulus and response. Influenced by Peirce's notion that *habit* was a powerful mediating force in sign-making processes, Morris developed a semiotic model that explained all sign processes in terms of *reactions* caused by the *presence* of signs under certain *conditions*. Hjelmslev was the theorist chosen by Peter Bøgh Andersen (Andersen 1990), when writing his *Theory of Computer Semiotics*, and Morris the one chosen by Heinz Zemanek (Zemanek 1966), when writing his pioneering *Semiotics and Programming Languages*.

### 25.2.3 Culture, signification and communication

Before we proceed to the next sections, let us revisit the saying that "meaning is in the mind of the beholder" for final remarks about conditions for human mutual understanding. We will borrow Umberto Eco's *Theory of Semiotics* (Eco 1976), which brings together Peircean and Saussurean elements in interesting ways. On the one hand, Eco adopts the notion that signs trigger an indefinitely long chain of other signs in the mind (he speaks of unlimited semiosis). On the other, he defines signs as content-expression correlates. The apparently diverging perspectives are reconciled with the introduction of *culture* in two fundamental semiotic processes: signification and communication.

*Signification* is a correlation between content and expression. *Signification systems* are collections of "discrete units [...] or vast portions of discourse, provided that the correlation has been previously posited by a social convention". *Communication* is a process in which "the possibilities provided by a signification system are exploited in order to physically produce expressions for many practical purposes." (Eco 1976: p. 4) Notice that Eco does not say that in communication one *repeats the conventions* encoded in signification systems. He says that communicators *exploit the possibilities provided by a signification system*. In other words, communication *may* include sign productions that deviate from the system, as well as others that don't. In fact, Eco's theory of sign production defines communication as a process that manipulates signs, "considering, or disregarding, the existing [socially conventionalized] codes" (Eco 1976: p. 152). The primacy of social conventions in communication, even when *stepping out* of existing signification systems, places *culture* at the center of Eco's Semiotics, which he actually characterizes as *the logic of culture*. Semiotics, in his view, is thus a comprehensive study of culturally-determined codes and sign productions at all levels of human experience.

## 25.3 SEMIOTICS AND HCI: TWO DISCIPLINES, TWO CULTURES

The work of Hirschheim, Klein and Lyytinen (Hirschheim et al 1995) on the conceptual and philosophical foundations of *Information Systems Development* (ISD) can further our understanding of why Semiotics and HCI have been keeping apart from each other. Following a still widely-accepted view that information systems are technical implementations of social systems, the authors trace the differences among ISD approaches, concepts, models, methods and even tools back to their proponents' interpretation of reality. Building on Burrell and Morgan's work in organizational analysis (Burrell and Morgan 1979), they list four *paradigms* in ISD research and professional practice. The paradigms are simplified world-view models produced by combining the end points of two axes: objectivity-subjectivity and order-conflict (see Table 25.1).

| Paradigm | Choice between Objectivity and Subjectivity | Choice between Order and Conflict | Assumptions about Reality |
|---|---|---|---|
| **Functionalism** | Objectivity | Order | There is an objective order in social reality, which can be known and described independently of subjective interpretation. |
| **Social Relativism** | Subjectivity | Order | Reality is a social construction, which is the result of interpretations determined by continuous cultural changes. |
| **Radical Structuralism** | Objectivity | Conflict | Reality is determined by objective super-structures that are in constant power conflicts. |

| | | | |
|---|---|---|---|
| **Neo-humanism** | Subjectivity | Conflict | Reality is continually transformed by multiple objective and subjective epistemologies that co-exist and contribute to historical evolution. |

Table 25.1: Hirschheim and colleagues' paradigms at the base of information systems development.

Different assumptions about reality, shown in the last column of Table 25.1, can be aligned with distinctions between Saussurean and Peircean perspectives on meaning. This is not to say, however, that the same set of paradigms can be generated by semiotic theories alone, and much less that Saussure or Peirce (would) have embraced the principles of functionalism, social relativism, radical structuralism or neo-humanism. We must only observe the fact that paradigms that preclude subjectivity have higher compatibility with definitions of meaning that do not speak about mental mediation than those that include it, and vice versa. Likewise, speaking only of the paradigms where *order* is important, we should observe that the defining characteristics of functionalism in Table 25.1 are more compatible with Cartesian canons than is the case with social-constructivism.

Hirschheim and co-authors note that most of the work in ISD published until 1995 tacitly falls into the functionalist perspective, according to which subjective interpretation doesn't play a role when systems requirements are elicited and data models are built, for example. Testing hypotheses against empirical data can, in this view, prevent subjective intuitions, beliefs and aspirations from contaminating scientific knowledge of an objective reality. Consequently, this paradigm has long supported a feedback loop in which information and communication technologies (ICT) are viewed and produced as the impersonal (*i. e.* non-subjective) result of applying empirically-tested facts and principles to solve problems. Computer scientists have thus been increasingly led into empirical research practices looking for more facts and principles, but speaking very little of their interpretations and valued perspectives regarding what these facts and principles *mean*, why, and under what kinds of conditions.

As a multi-disciplinary field, HCI encompasses almost as much diversity in terms of scientific traditions as the number of disciplines that can be listed among its contributors (*e. g*. Psychology, Anthropology, Sociology, Design, Ergonomics and Computer Science, to name but only a few). However, most of its success and prestige is associated to ever more sophisticated and pervasive kinds of interactive products coming from the ICT industry. In this context, *predictive knowledge*, based on sound empirical experimentation, is in high demand. The industry is avidly looking for tested facts and principles to guide its engineering processes and found the establishment of norms and standards. Thus, the pressure of industrial demands on HCI research has been reinforcing functional perspectives on science and restating the belief that scientists and technologists do their work *impersonally*, backed by statistical inferencing methods applied to objectively collected empirical data.

Semiotics has been successfully applied in marketing and advertising (Umiker-Sebeok 1988; Floch 2001), for example, which suggests that HCI and Semiotics could work together to respond to industrial demands. Nonetheless, the power of opposing *scientific* traditions and practices has been keeping them on separate tracks. Semioticians of Peircean extraction, for instance, are interested in the rich variety of meanings that emerge in sign production and sign interpretation contexts. They may use extensive interpretive and hermeneutic methods to generate in-depth understanding of situated meaning-related processes, before they engage in statistical inferencing methods leading to general signification facts and principles. In addition, they must not dismiss the fact that in order to establish the very object of their investigation (a sign or sign process) they need the mediation of *their* own minds. That is, in order to take anything as a *sign* (or evidence) of other signs or sign processes of interest, investigators must carry out *interpretation*. The process and product of their investigation will depend totally on other signs: their inferences, their validation criteria, their conclusions, and last but not least on the way they *signify* their knowledge discovery process and communicate it to the scientific community. The main criteria for scientific validity, which must be made apparent in this specific situation, is that newly discovered signs can: be framed into coherent inferential discourse that consistently relates to previously existing knowledge; correct, expand or innovate aspects of previous scientific findings; be exposed to the evaluation of other researchers in the scientific community; and produce, as a result of such exposure, the advancement of collective scientific knowledge.

This view contrasts in important ways with the one that credits scientific validity only to research results obtained by an empirical testing of hypotheses and subsequent statistical inferencing. Firstly, although a semiotic view acknowledges the scientific validity of knowledge produced with these methods, it requires that the choice of hypotheses be accounted for as a *signification process*. Why have the hypotheses been chosen? Why are they significant *per se*, or more significant than other possible hypotheses? Secondly, and by consequence, it brings into the territory of legitimate *scientific activity* the very elaboration of candidate theories and hypotheses that will be tested in the course of empirical knowledge discovery processes. By means of rigorous interpretive methods, Semiotics is ready to inspect and validate the processes by which scientists assign meaning to (*i. e.* “interpret”) their hypotheses, their testing and their results, against the backdrop of theories that they propose to be building.

These contrastive characteristics can explain why semiotic knowledge is usually viewed by the HCI community as speculative and subjective, an impractical philosophic discussion that cannot usefully inform the design and development of ICT products. By the same token, they can also explain why researchers with conscious or unconscious semiotic awareness tend to be skeptical of HCI theories built upon hypotheses whose selection is not carefully and consistently accounted for. Moreover, motivated by the huge diversity of human meaning-assigning strategies, these researchers may be far more attracted to the *possible* variations of users' interpretive behavior in view of even small contextual changes than to regularities found under highly controlled tests in laboratory.

Given the above-mentioned differences in purpose, practice, beliefs and values, the most difficult obstacle for bringing together Semiotics and HCI is probably not the fact that Semiotics requires extensive learning of unfamiliar concepts and methods. This has always been the case in the multidisciplinary research practices that characterize HCI. The real challenge for the two disciplines seems to be how to combine different epistemologies and rise above scientific validity disputes. Only then can Semiotics seed progress in HCI.

The rest of this chapter is my personal narrative of a successful case, the theoretical construction of Semiotic Engineering (de Souza 2005; de Souza and Leitão 2009). I hope many others are on their way to correct and refine our thinking, partaking in the ongoing scientific *semiosis*.

## 25.4 COMPUTERS AS MEDIA

Today, viewing *computers as media* is like viewing cars as vehicles. Of course, computers enable, support and enrich individual communication, group communication, and mass communication. We can even put it the other way around: All contemporary *media* involve computation. But, how obvious was this connection when HCI emerged as a discipline? And, how obvious right now are the consequences of this view for current and future computation?

Viewing *computers as media* is the hallmark of all semiotic approaches to computing and human-computer interaction. In 1988, Kammersgaard proposed that human-computer interaction could be characterized in substantially different ways (Kammersgaard 1988). One of them was called *the media perspective*, which involved not only the fact that computers could be used by people to communicate with each other, but also the fact that computer application programmers could communicate with users through systems interfaces. As the author acknowledges, the latter was influenced by illustrious predecessors:

> "*I will not go into further detail about this last type of communication, except to mention that Oberquelle, Kupka & Maass (1983) talk about delegation of communicating behaviour from the designer to the machine and then treat the situation as seen from a dialogue partner perspective, whereas Andersen (1985) treats the designer as having the role of one sender in a collective of senders, who makes a contribution to each message sent through the medium.*"
>
> *-- Kammersgaard 1988: p. 356*

Only a few years before, Winograd and Flores had published a notorious book, in which they proposed to shift the focus in computing from pursuing ever-increasing artificial cognitive capacities to supporting pervasive human communication and coordination processes (Winograd and Flores 1986). And in 1990, Andersen would publish the first comprehensive account of *computers as media* in his *Theory of Computer Semiotics* (Andersen 1990).

From the late 1980's to the mid 1990's, a number of researchers, in different parts of the world, started thinking about how to articulate Semiotics and Computing. In 1988, in a 55-page chapter in Hartson's *Advances in Human-Computer Interaction* (Hartson 1988), Nadin (Nadin 1988a) explored the semiotic implications of interface design and evaluation. In the same year, he would also publish an article in *Semiotica*, about a semiotic paradigm to systems interface design (Nadin 1988b). In 1993, Andersen edited a book explicitly called *The Computer as Medium* (Andersen et al 1993), and we published our first paper on the semiotic engineering of human-computer interfaces (de Souza 1993).

All of this early work explored the power of viewing computer programs and systems — and especially their interfaces — as *signs* of a very specific kind. They are designed and engineered according to the laws and limits of computation, but what they ultimately communicate is a message meant *by* humans (systems designers and developers) and *to* humans (systems users). This is the rationale for viewing human-computer interaction as computer-mediated human communication, the powerful idea underlying Winograd and Flores's *manifesto* for a language-action perspective (LAP). LAP attracted many semiotically-inspired researchers in Computer Science (Communications of the ACM, May 2006). Most of them turned towards Organizational Semiotics (Liu 2000), which in general is closer to information systems development than to human-computer interaction design. In Brazil, however, a group of researchers has been exploring the use of Organizational Semiotics concepts in the design of human-computer interaction (Baranauskas et al 2003; Bonacin et al 2004; Neris et al 2011).

Semiotics was also implicit or explicit in the work of other researchers working in HCI. For example, Mullet and Sano (Mullet and Sano 1995) explored semiotic features of representations used in the design of visual interfaces. Their emphasis was on how to *communicate* the design intent to users through interface signs, an idea that was gaining momentum at the time. In fact, computer systems interfaces provided an excellent context to revisit the concept that there are *languages* of/in design. The chapter by Rheinfrank and Evenson in Winograd's *Bringing design to software* (Winograd 1996) is a great illustration. The authors discuss *design as communication*, whether intentional or not, and advance the idea that when it is intentional, the use of *design languages* can increase the effectiveness of communication.

As computers became more pervasive and took over a larger portion of social processes, the *computers as media* perspective gained the interest of researchers investigating the users' response to intentional communication by service providers (Light 2001) and information appliances in general (Fogg 2002), for example. Fogg's work, for one, developed into an elaborate study of persuasion through technology (Fogg 2003), which has been interestingly explored in the context of games and culture-sensitive communication (Khaled et al 2006). In a recent work, O'Neill (O'Neill 2005) theorizes about interactive media in view of contemporary technologies that provide new mediation opportunities for human semiotic processes. Research on the convergence of Semiotics, interactive technology and literacy is also noteworthy. Marion Walton's work (Walton 2008) in South Africa provides insightful elements to the study of *computers as media*, enriched by empirical evidence collected in African school children's encounters with computers and the Internet (Prinsloo and Walton 2008).

Inspired by Andersen's pioneering work in Computer Semiotics, we developed an extensive semiotic account of human-computer interaction, Semiotic Engineering (de Souza 2005). Having started as a semiotic *approach* to user interface language design (de Souza 1993), over the years Semiotic Engineering evolved into a semiotic *theory* of HCI with its own ontology and specifically-designed methods of investigation (de Souza and Leitão 2009). This is the theory I will use in the rest of this chapter to illustrate, in practice, the kinds of connections existing between Semiotics and HCI.

### 25.4.1 Semiotic Engineering

Semiotic Engineering picked up the early view that human-computer interaction is in fact computer-mediated human communication. It then defined and articulated a number of fundamental concepts, their relations with each other and their implications not only for the development of the theory itself (the internal motivation), but as a contribution to the advancement of HCI (the external motivation). The most striking distinction proposed by Semiotic Engineering compared to other theories and conceptual frameworks in HCI (Carroll 2003) is to develop the early suggestion put forth by Kammersgaard (Kammersgaard 1988) and postulate that designers of interactive software are active participants in the communication that takes place through user interfaces. They communicate their design vision to users by means of interface signs like words, icons, graphical layout, sounds, and interface controls like buttons, links, and dropdown lists. Users unfold and interpret this message as they interact with the system. In other words, the communication between users and systems is in fact part of a *meta*communication process; part of the communication process initiated by designers about how, when, where and why to communicate with the system they have designed.

Since designers cannot be personally present when a user interacts with software, they have to *represent* themselves in the interface, using a specifically designed signification system, and subsequently *tell the users* what the software does, how it can be used, why, and so on. The choice of representations is actually wide. For instance, designers may represent themselves as humanoids (*e. g.* systems interfaces with human characteristics like affect and natural language abilities), as machines (*e. g.* systems interfaces with press buttons, slide controls, dials, and the like), or even as spaces (*e. g.* systems interfaces with virtual worlds that users explore to achieve various kinds of effects). Depending on the message that they have to communicate, some representations will work better than others. Humanoids, for example, are likely to communicate explanations and instructions more easily than virtual spaces. Machines, in turn, convey physical affordances more efficiently than the natural language discourse of humanoids.

No matter their *self*-representation, through systems interfaces designers are *telling* users:

- what they know about the users (who they are, what they know, what they wish or need to do, in which preferred ways, and why);
- how they have responded to the users' needs or aspirations (what system they have built and how it works)
- what values are encoded in their response (why and how does the system improve the users' lives).

Here is a very simple sketched example of how the elements above are communicated by designers through the interface. We contrast Microsoft Word (Figure 25.5) and Open Office Writer (Figure 25.6) **tools** menu, showing how the designers' message comes across.

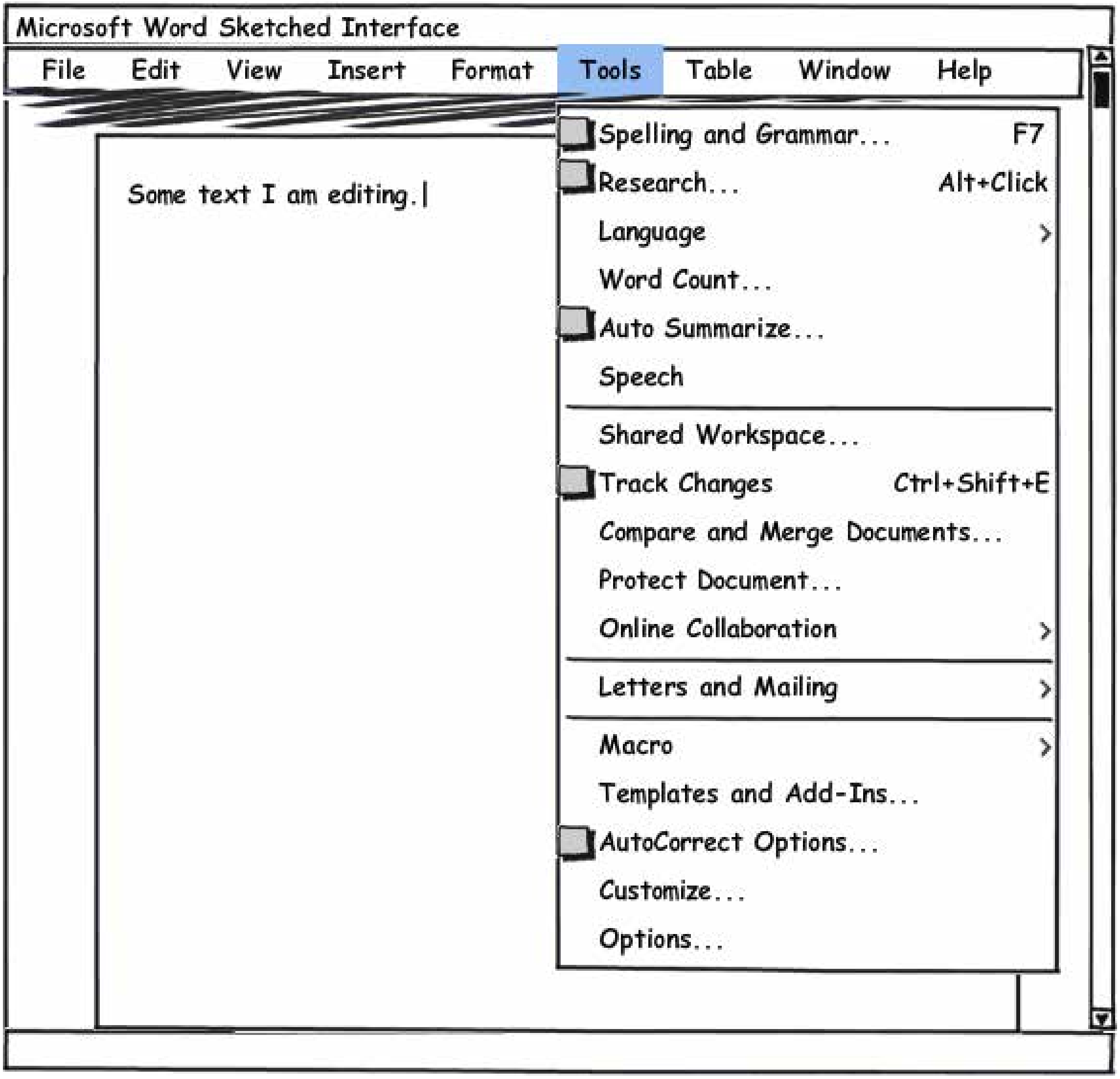



Figure 25.5: A sketched version of Microsoft's Word interface unfolding the Tools menu.

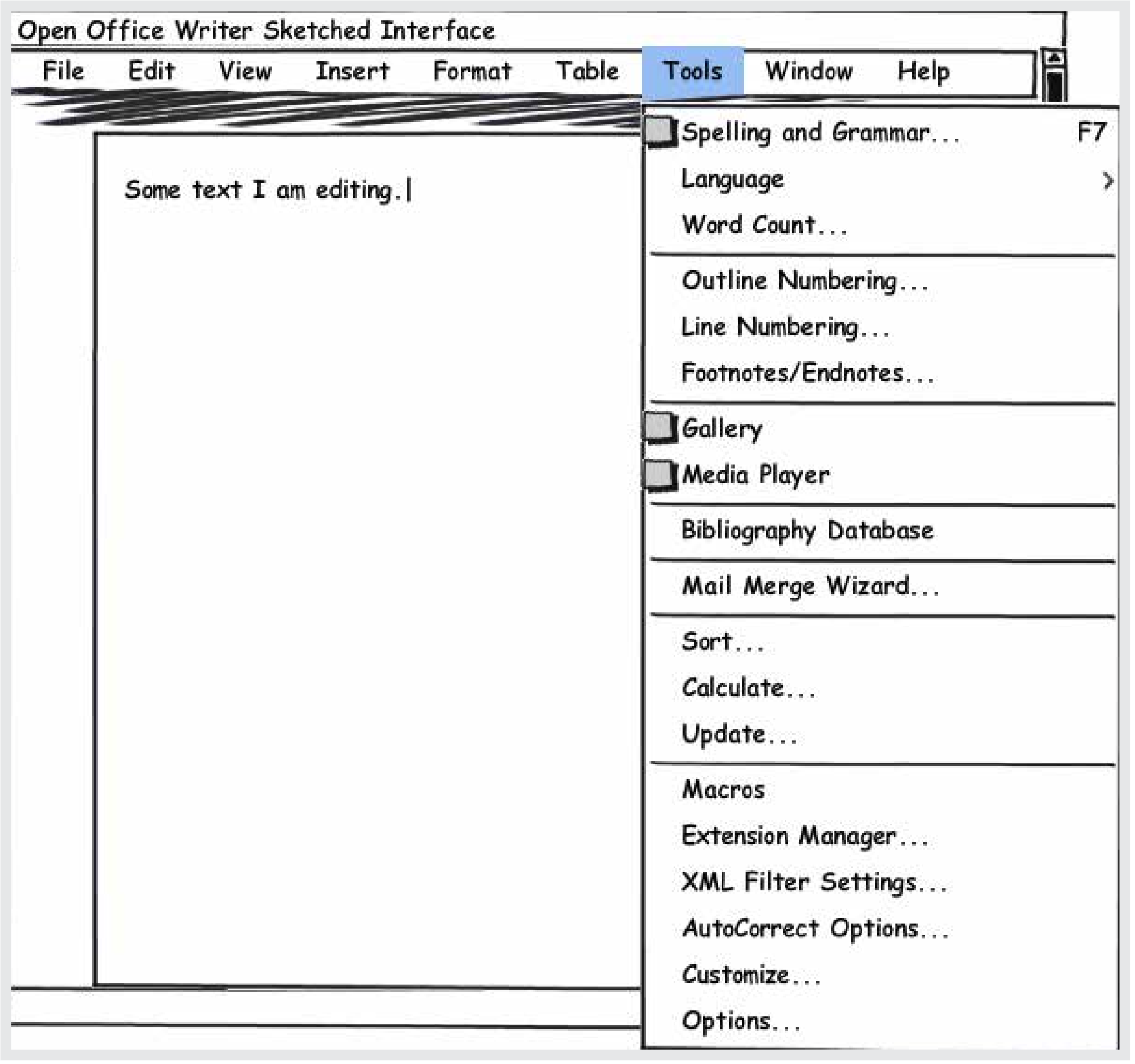
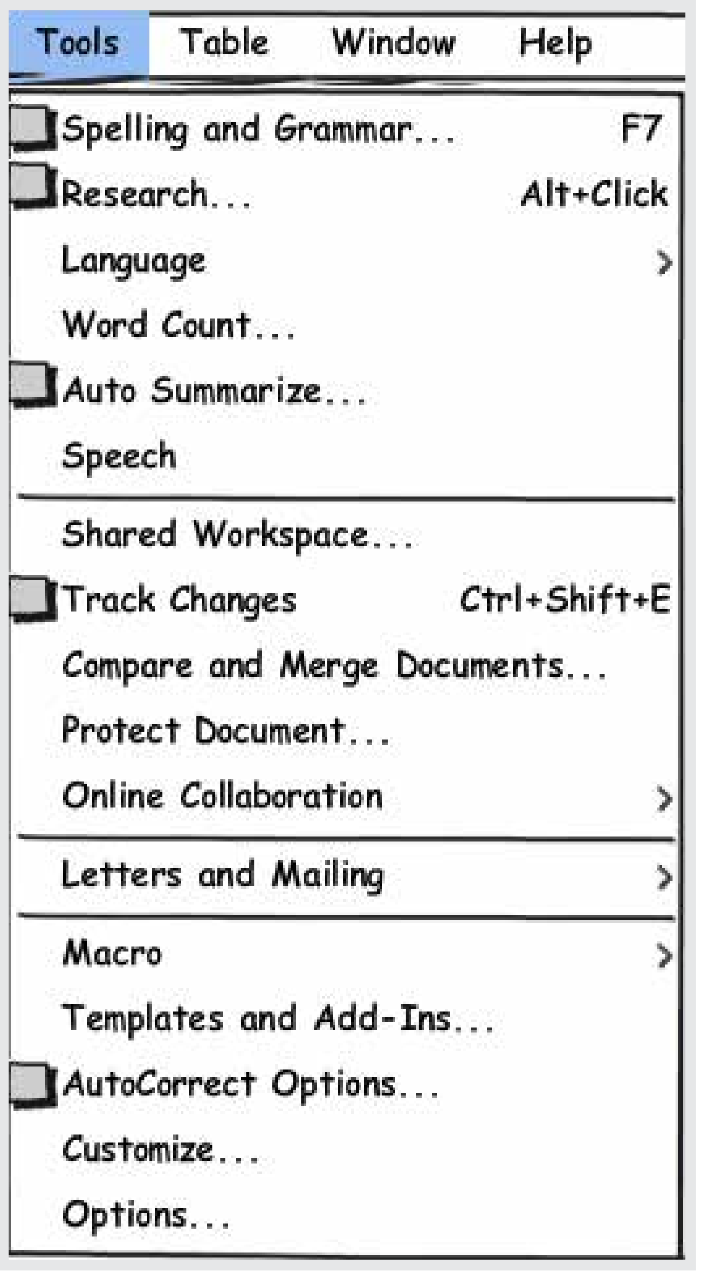



FIGURE 25.6: A sketched version of Open Office's Writer interface unfolding the Tools menu.



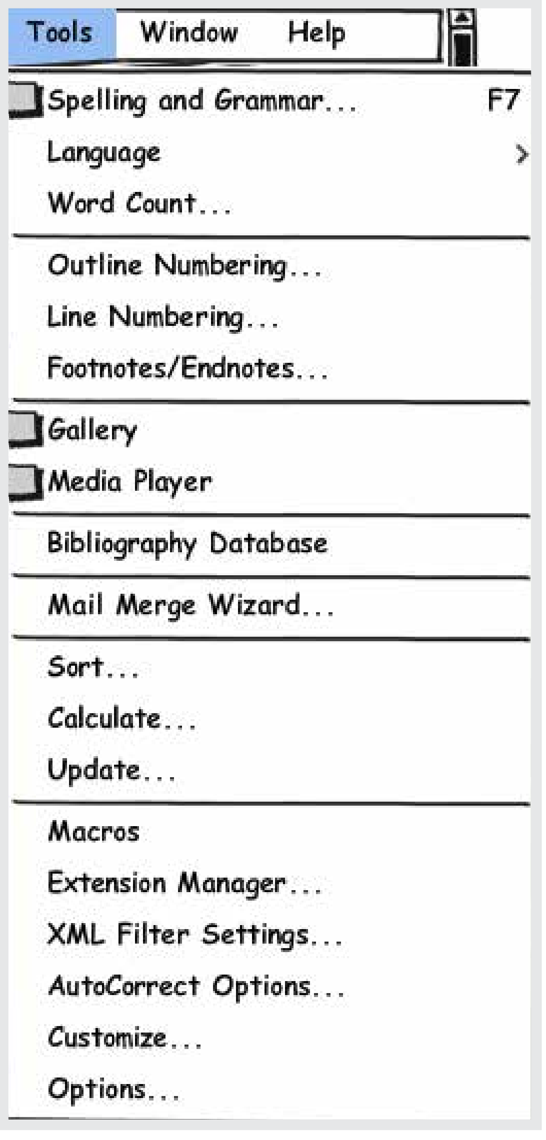
The contrast of both designs shows that they communicate different views on users' needs and opportunities. The following is a summary of the most striking differences between the two messages communicated through Microsoft Word and Open Office Writer interfaces.

### 25.4.2 Contrast between MS Word and the OO Writer interfaces



## Microsoft Word

MS Word Tools can be grouped into 4 categories (list items separated by a line): the first with tools related to document content; the second to document manipulation and authoring; the third to mail preparation; and the fourth to interface customization and extensions.



## Open Office Writer

OO Writer Tools can be grouped into 7 categories (list items separated by a line): the first with tools related to document content; the second to document structure; the third to inclusion of media files; the fourth to bibliography handling; the fifth to mail preparation; the sixth to data field manipulation; and the seventh to interface customization and extension.

The designers of both applications are telling their users that tools applying to document content are *top of the list*, and that 'Spelling and Grammar' tools are *the* most important ones. By comparison, we (like users) *get the message* that MS Word provides more resources than OO Writer in this particular category. In fact, OO Writer tools are a subset of MS Word's. Similarities in shortcuts are remarkable, suggesting that OO Writer's designers expect their users to have (or to have had) experience with MS Word at some point in their lives.

Similarities extend also to other categories. In both applications, there are tools to handle mail preparation and to support customization and extensions. The tools in these categories, however, may have different names, and OO Writer provides an additional tool, compared to MS Word: it lets the users customize 'XML Filters Settings'.

At a closer look, we see that differences in names actually communicate a different perspective on the tools themselves. Whereas in MS Word names tend to express *objects* (with the exception of 'customize'), in OO Writer the designers' communication, when different from MS Word's, alludes to *agents*: 'Mail Merge Wizard' and 'Extension Manager'. In these two cases, the communication is clearer (more precise) than in MS Word, which may achieve the effect of encouraging novice users into experimenting new features (since a *wizard* and a *manager* will be there to help them).

| **Microsoft Word** | **Open Office Writer** |
|---|---|
| MS Word's designers communicate very clearly their emphasis on group collaboration. They are *telling* users that document preparation is or can be a group activity, requiring such things as shared workspaces and online communication, in addition to tracking changes and comparing, merging and protecting documents. The communication of their design vision at this point clearly locates users in a computer-supported organizational context. | OO Writer's designers, by contrast, do not talk about computer-supported collaboration. The communication of their design vision at this point portrays an individual user, working on the details of his/her document. The designers give special emphasis to document structure (numbering and notes) and bibliographic reference management, communicating that they want to support users in preparing complex extensive documents. |

By comparison with OO Writer's designers' communication, MS Word's designers communicate their concern with usability in different ways. Five of their tools come with icon representations that help users identify their presence in toolbars (a more concise and more easily accessible interface control than menus). Three menu options have associated keyboard shortcuts, and four options unfold into submenus, rather than invoking dialog windows (communicated by '…'). Together these *messages* tell the users that MS Word's designers are concerned with providing fast access and fast learning support for their users. In contrast, OO Writer's designers favor longer dialogs ('…'), which actually is in line with longer processes that can also be expected from interaction with a 'wizard' and a 'manager'. Furthermore, by communicating that they expect their users to need a bibliographic database, OO Writer's designers give us the impression that they have worked for meticulous users engaged in longer-term tasks. By contrast, MS Word designers give us the impression that they have worked for busy users, engaged in broader-context activities of which document preparation is only a part.

In both cases, however, the *complete* message about which tools are available and how valuable they are expected to be for the targeted users is not likely to be understood unless users engage into interacting with the application. In MS Word, for example, in order to understand what "research" really means in context, users will probably click on the option (*i. e.* interact with the message itself) and see what comes next. If they do, they'll find out that the designers are offering the possibility to connect to and search online reference books and other resources. Likewise, in order to understand what "bibliography database" is about in OO Writer, users will also probably click on the option. If they do, they will find out that this option connects them to an existing 'bibliography' that is installed with OO Writer, and which can be extended or modified to accommodate the users' reference items. In both applications, the users can resort to online help resources to learn more about these and other parts of the designers' message conveyed in the menus above. Thus, we see that communication is not limited to visible signs in isolated screens — communication is also conveyed in the process of subsequent communication with the application and its resources (*i.e.* online help and documentation).

The summative contrast between communication expressed by the MS Word and the OO Writer interfaces is a good sampler of the essence of Semiotic Engineering, whose main tenets are the following:

1. HCI is an instance of *metacommunication* — communication about how, when, where and why to communicate with computer systems. The metacommunication message content can be summarized as follows (note the presence of the designers using 1st person pronouns like "I", "my", etc.)

*"Here is my understanding of who you are, what I've learned you want or need to do, in which preferred ways, and why. This is the system that I have therefore designed for you, and this is the way you can or should use it in order to fulfill a range of purposes that fall within this vision."*

2. Both designers and users are engaged in metacommunication; the system's interface represents its designers at interaction time, it is the designers' deputy. The interface enables all and only the designed types of user-system conversations encoded in the underlying computer programs at development time. The metacommunication message from the designers is unfolded and received as users interact with it and learn 'what the system means'.

3. There are three classes of metacommunication signs that designers can use to communicate their message to users. *Static signs* communicate the essence of their meaning with instant time-independent representations (typically representations that can be correctly interpreted in static screen shots). *Dynamic signs* communicate the essence of their meaning with a series of time-dependent representations (typically representations that can only be correctly interpreted over a number of subsequent screens or states of the system). *Metalinguistic signs* are static or dynamic signs that differ from either the former or the latter because the essence of their meaning is an explanation, a description, an illustration, a demonstration or an indication *of other [interface] signs* (typically textual or video material referring to the meaning of some other static or dynamic sign).

4. Metacommunication signs in HCI must be produced computationally. Consequently, this specific kind of computer-mediated human communication introduces critical constraints in sign production processes. Systems designers must create representations that by necessity have a single definitive encoded meaning — no matter if the designers (and the users) can easily produce evolved meanings for these representations in natural sign-exchange situations. The algorithmic nature of the *medium* in which metacommunication takes place mechanizes human semiosis, in both directions (designers' signs produced *by* the system interface and users' signs produced *with* the system interface).

5. The quality required for metacommunication to be efficient and effective is *communicability*, a system's ability to signify and communicate the designers' intent (which is ultimately to satisfy the users). The evaluation of communicability involves a methodical analysis of how the designers' message is emitted (composed and sent through the interface) and of how it is received (interpreted and followed by physical and/or mental sign-mediated action) by the users.

6. Because the users' response to the designers' metacommunication must be mediated by interface signs, the users must *learn* the interface language, a unique signification system, in which the designers' message is fully encoded. Users learn this language in the very process of using it in interaction. This process is similar to natural language acquisition, which humans are fully equipped to do, *except* that the users' immediate interlocutor (unlike humans) does not reason abductively. Therefore, some of the users' communicative intent may persistently fail to be interpreted simply because the designers have not anticipated the users' sign-making strategies. Because this can cause disorientation in the process of interface language acquisition, designers should also ex-

plicitly signify the communication principles that they have chosen to encode their message.

Semiotic Engineering has been gradually making its way into HCI, especially after Don Norman, one of the leading figures in early UCD (Norman 1986), wrote about advantages that he sees in a semiotic engineering perspective on HCI (Norman 2004; Norman 2007; Norman 2009). In the next section, we begin to conclude this chapter with very brief concrete examples of how this perspective influences the analysis and design of computer-mediated communication in HCI.

## 25.5 A CLOSER LOOK AT COMPUTER-MEDIATED COMMUNICATION

As mentioned earlier in this chapter, saying that computers are *media* has become a cliché in the turn of the century. Computers are everywhere, computing is embedded in hundreds of things that we carry and encounter in daily life, and digital information exchanged through computers, with or without our control, affects almost every step we take as individuals and as a society. A relevant question for HCI is: How well are HCI *theories* prepared to inform and equip HCI designers in this context?

The question is actually not new. In 1986, Winograd and Flores asked it from the Artificial Intelligence community, mainly, hitting additional targets in HCI and in Computer-Supported Collaborative Work (Winograd and Flores 1986). In spite of this history, however, most of the existing HCI *theories* are dominated by cognitive perspectives. They can and do inform the design of interaction with respect to facilitating the learning, memorization and retrieval of productive interactive patterns, for instance. And nobody disputes that this is a fundamental requirement in HCI design. However, the *user-centered* approach, as its name so clearly 'signifies', intentionally or unintentionally has treated *design meanings* as if they were mind-independent entities that are *somehow* elicited from users and then reified in design models and prototypes before they eventually take the shape of a system's interface. The designers' interpretation and signification processes are *not* accounted for in original UCD (Norman 1986), whose theorists have in general tended to follow the Cartesian tradition of postulating the existence of *primary* cognitions. These are shared by all humans (and thus by all users), which stimulates research seeking for universal primitives and, based on their consequences, making predictions about users' behavior.

The problem of UCD in the *computer as medium* paradigm is that it cannot accommodate computer-mediated communication between systems designers and users. Designers, quite plainly, do not belong in UCD interaction models. Neither of the two most influential historical sources of UCD — the seven-step theory of action (Norman 1986) and the human information processing model (Card et al 1983) — account for the *designers'* cognitive processes and meaning-making activity. This part of the story has been covered by methods (*e. g.* ethnography (Bouissac 1998) and theories (*e. g.* Activity Theory (Kaptelinin and Nardi 2006)), which have nonetheless failed to form with UCD a seamless body of knowledge that can satisfactorily account for the whole design process in accordance with the *computer as medium* perspective. The final *ad hoc* combination of uncohesive parts in actual design processes seems to be what Spool refers to as the "skills and talents of practitioners" (p. 8, above).

Although the new *media* perspective instantly reminds us of Web applications design and of interaction with or through mobile devices, there is yet more to it. Digital literacy, which is undoubtfully a requirement for the achievement of full citizenship in the 21st century, has evolved to include computational thinking skills without which users (citizens) are not likely to be able to engage in the new *cultures of participation* (Fischer 2011).

In the abstract of a thoughtful and provocative article, Jeannette Wing (Wing 2008) expresses her perception of the extent to which changes in current computing practices and possibilities affect science, technology and society:

> "*Computational thinking will influence everyone in every field of endeavour. This vision poses a new educational challenge for our society, especially for our children. In thinking about computing, we need to be attuned to the three drivers of our field: science, technology and society. Accelerating technological advances and monumental societal demands force us to revisit the most basic scientific questions of computing.*"
>
> -- *Wing 2008: p. 3717*

In the following, I will present three contemporary examples of different levels of computational thinking involved in new kinds of social interactions. In each case, I will underline the potential advantages of a semiotic framing for the illustrated phenomenon. Then, in a separate sub-section, I will talk about research we are doing at SERG, the Semiotic Engineering Research Group of *Pontifícia Universidade Católica do Rio de Janeiro* (PUC-Rio), partnering with colleagues in Colorado University at Boulder and *Universidade Federal Fluminense* (UFF), a public university in the State of Rio de Janeiro, Brazil.

### 25.5.1 Some contemporary examples

In the illustrations below, we will see how computer-mediated human communication may involve explicit computational representations of *self* and *message* at different levels of complexity. In each case, the *purpose* of communication is very prosaic — to help someone use his email account. The innovation lies in the chosen *form* of communication and the wealth of new meanings that it introduces in human experience. We will also see that computing literacy is the key factor for reaping the benefits of this new means of self-expression and social participation. Taking a semiotic perspective on literacy — the ability to interpret and produce signs of socially-valued signification systems in order to achieve social participation and full citizenship — the point of the illustrated cases is to show that being able to program computers has become as important in the 21st century as reading, writing and counting has been in previous centuries.

#### 25.5.1.1 Example 1: Programming via parameter setting

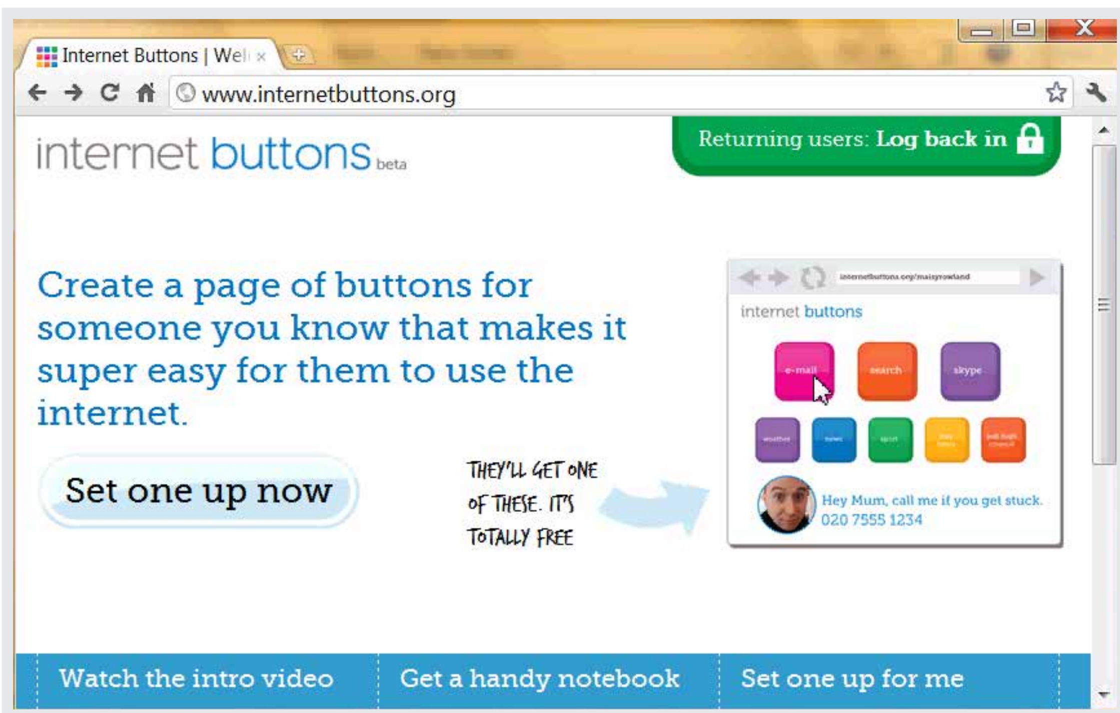



Figure 25.7: Internet Buttons” entry page.



The first example is **Internet Buttons®**, a tool that allows you to create a personalized page with buttons that direct you straight to your favorite web sites, services and applications. **Internet Buttons** was created by a not-for-profit company in the UK "to get people from different generations talking more, sharing more and spending more time together." In Figure 25.7, we show their home page. Notice how they *get their message across* with different kinds of signs like text ("Create a page of buttons for someone you know [...]" and "Hey Mum, call me if you get stuck [...]") and image (colorful buttons and personal picture) and think of how obviously the software designers not only participate actively in interaction with the user, but also of how they would completely fail to achieve their intent if they did not have the ability (a "personal talent"?) to express *themselves* through software.

The beauty of **Internet Buttons** lies, however, in the recursive nature of the designers' message. Through their program, they are inviting Internet users to *program* "super easy" Internet interaction for "someone they know". The programming paradigm is also an extremely simple form of parametric procedure where all that the users have to do is provide the correct values for pre-selected parameters. Therefore, the kind of computational thinking required is very basic. In Figure 25.8, we show a step of the *programming* required to access a Gmail account.

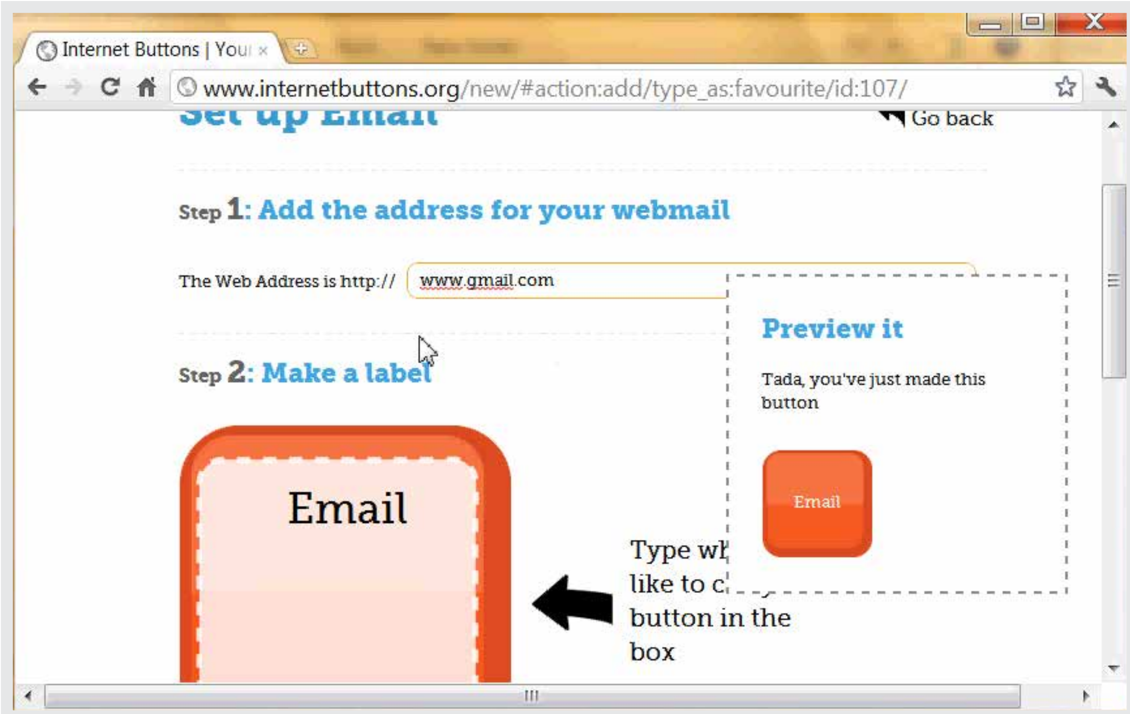



Figure 25.8: Creating an Internet Button to access Google's Gmail login page.

*Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.*

As the **Internet Buttons** entry page shows, once the buttons are *programmed* they go into a personalized web page where the user can place any number of buttons especially created for someone else. The personalized page template allows the user-programmer to add a personal picture and a message directly addressed to whoever he are talking to through **software that speaks for him**. Therefore, the software carries a representation of "self" in addition to a representation of the "message" to be communicated.

#### 25.5.1.2 Example 2: Programming via macro recording

Our next example is IBM's **CoScripter®**, a macro recorder for the Web. Unlike **Internet Buttons**, **CoScripter** does not focus on *personal* communication and representations of *self*. However, just like **Internet Buttons**, the purpose of **CoScripter** is to make Internet processes easier — for the user himself/herself and for whoever needs help with Web interaction. The system allows the user to *record* and *playback* interaction steps, optionally using the values of variables stored in a "Personal Database" in the user's machine. Recorded *CoScripts* are stored in an Internet server and can be shared with other users. In Figure 25.9, the arrows that we added to the snapshot show how, in this case, the CoScript instructions (on the left) use the information in the Personal Database (at the bottom left) to command interaction with the Gmail login page.

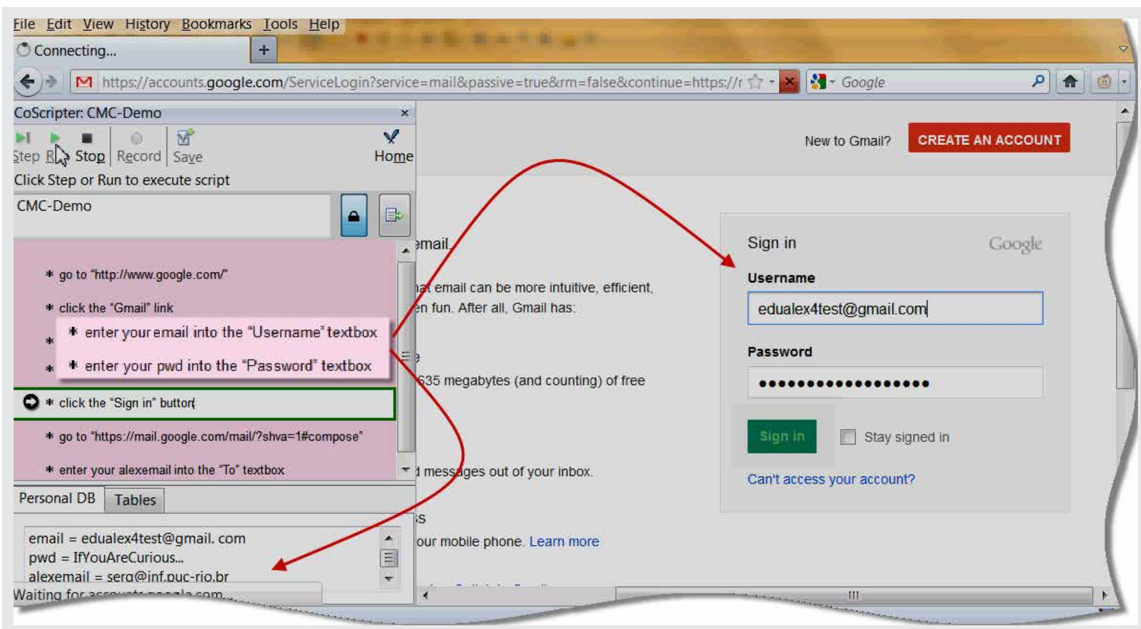



FIGURE 25.9: The execution of a CoScript for accessing a Gmail account.

*Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.*

Although the *scripters* are not explicitly represented in the CoScript (an interesting sign of the effect of *impersonal* and *implied* communication between script designer and script user), the purpose of **CoScripter** creators is also interpersonal communication through software. In their home page, visited in December of 2011, we read the following message:

*"CoScripter is a system for recording, automating, and sharing processes performed in a web browser such as printing photos online, requesting a vacation hold for postal mail, or checking flight arrival times. Instructions for processes are recorded and stored in easy-to-read text here on the CoScripter web site, so anyone can make use of them. If you are having trouble with a web-based process, check to see if someone has written a CoScript for it!"*

*-- http: //coscripter.researchlabs.ibm.com/coscripter*

The programming paradigm in **CoScripter** is more sophisticated than in **Internet Buttons**. One of its creators, Allen Cypher, is a leading figure in programming by demonstration (Cypher 1993). The user-programmer in this case is dealing with a simple programming language, in which demonstrated interactions are automatically encoded. Optionally, the user may manipulate *variables* to make CoScripts more general and reusable in similar, but not identical, contexts. Mainly because of variable manipulations, the level of computational thinking required to use CoScripter is intermediary. There are a number of interesting semiotic issues to explore with **CoScripter**. In this chapter, I will only briefly remark that, unlike **Internet Buttons**, whose messages are computationally encoded with fixed "token-level" semantics (each button *means* a single Internet address), **CoScripter** messages can be computationally encoded with fixed "type-level" semantics (each script can be executed with different *parameters* specified in the end user's Personal Database). Each CoScript *means* a range of possible interactions with the same web page, web service or web application. All of the interactions are predicated by the same interactive *steps*, but there can be variations in *contextual parameters*. Thus, the communication content that can be expressed with **CoScripter** is considerably more complex (and powerful) than in the previous example. The interested reader can see a deeper semiotic analysis of **CoScripter** in one of our previous publications (de Souza and Cypher 2008).

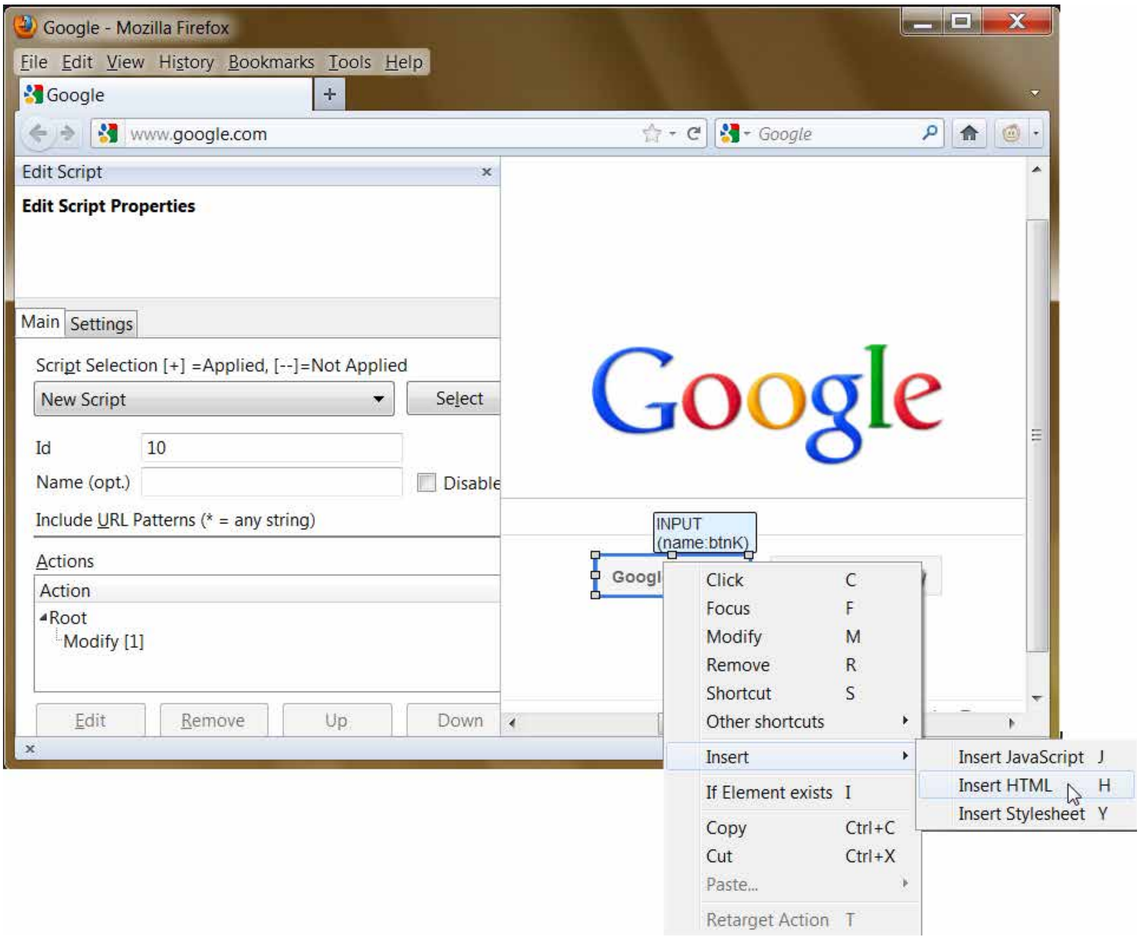



Figure 25.10: Programming with "Customize Your Web"

*Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.*

### 25.5.1.3 Example 3: Programming on wheels

Our last example is **Customize Your Web**, a personalization tool that allows users to change the appearance of web pages and add new functionality to them without having to dive too deeply into JavaScript programming if they don't want or don't know how to do it. **Customize Your Web** is an extension to Firefox designed by Rudolf Noe. In Figure 25.10, we show the end user programming environment offered by this tool. The user can select elements of a Web page (in this example, the "Google Search" button is selected) and make changes to its appearance and behavior (see the menu from which the user is about to select between "Insert JavaScript", "Insert HTML" or "Insert CSS"). Interface elements can be deleted, inserted and relocated as desired, as long as they can be *uniquely identified* on the existing page whenever it is loaded.

Compared to the previous example, **Customize Your Web** opens even more powerful possibilities for computer-mediated communication. In Figure 25.11 and Figure 25.12, we show how facilitating interaction with Gmail is substantially different in **Customize Your Web**, compared to **CoScripter** and **Internet Buttons**. Because the programming paradigm is close to *programming on wheels*, the possibilities for representation of *self* and *message* are limited only by technicalities of the tool (like problems with unique identification of elements on certain Web pages) and the resources of the user. The level of computational thinking required to use this tool effectively is advanced.

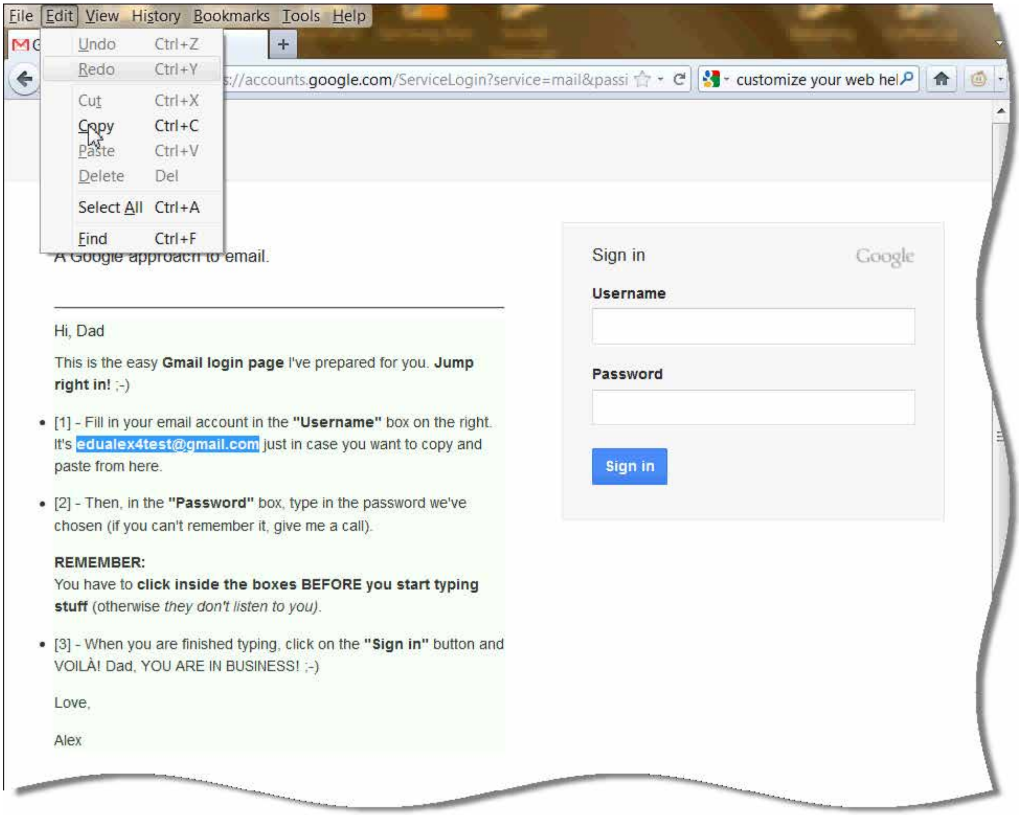



Figure 25.11: Gmail login page modified with “Customize your Web”



In Figure 25.11 (see customized text starting with “Hi, Dad” on the left), we can see that the user can *script* his own *presence* and *participation* in someone else’s experience with Gmail. The complexity of self-representation and message in this case is considerably high. Notice that in Figure 25.12, the *conversation* between the scripter and the targeted user (his Dad) extends over whole interaction spans with Gmail. In other words, a parallel *communication about communication* with Gmail (*i. e.* genuine metacommunication about Gmail) is **explicitly** and **intentionally** in place. This is one of the best examples of why the *computer as media* perspective calls for a different breed of HCI theories in order to help end users take the best out of the virtually unlimited possibilities of social interaction and participation now available for them.

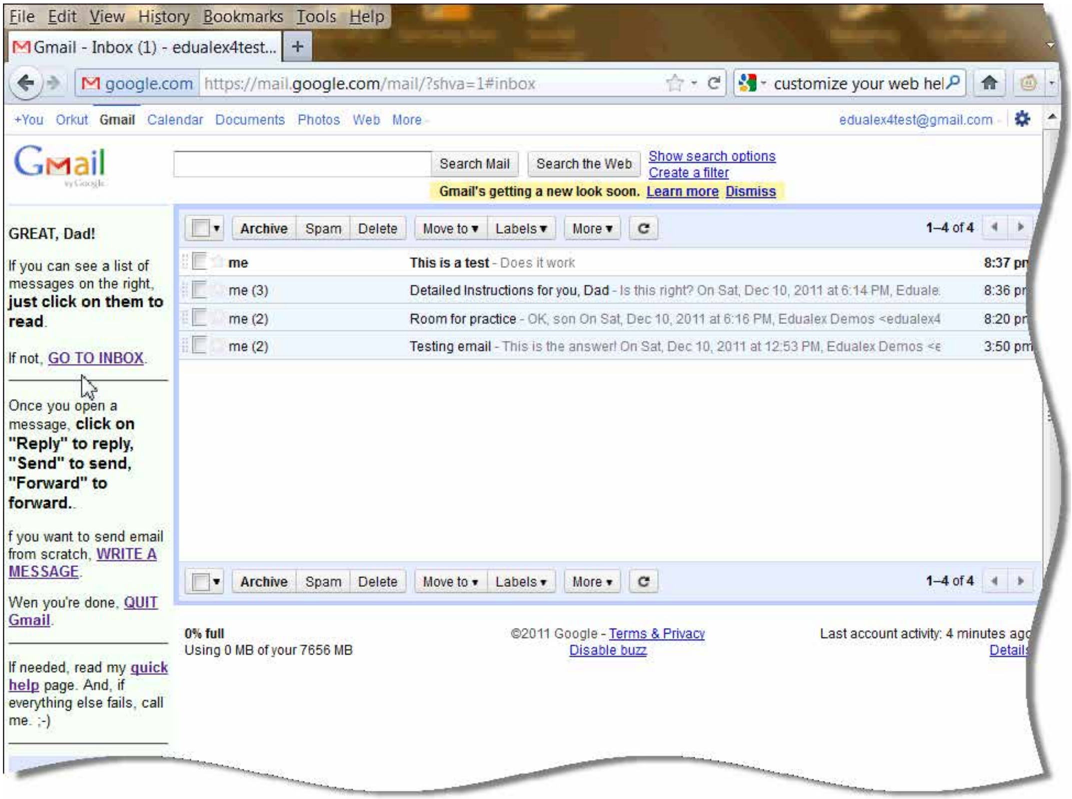



Figure 25.12: Gmail inbox page modified with "Customize your Web"



Designing cohesive and consistent dialog turns throughout a Web process, like shown in the **Customize Your Web** example, is *end user HCI design* — not only end user programming and development. However, theories that do not include *the designers* in the communication process that they are designing have very little to say about how (and why) the current interface of **Customize Your Web** (see Figure 25.10) must be improved. For example, UCD is not going to help the user *design all* the conversational paths that his parallel communication with Dad might take *in his absence*. Dad can decide to use any of the available interface controls during interaction with Gmail. So, how does the user build a representation of *conversational context* that is consistent with what he is telling Dad at each step of the programmed interaction?

The answer to these questions can not only show the complexity of programming that end users can engage in at this stage of technological development, but also — and more importantly — it can show what kinds of new dimensions a semiotic perspective can bring to HCI research and practice. In the next sub-section, we will very quickly mention how we are exploring computer-mediated communication issues with Semiotic Engineering.

### 25.5.2 Going a few steps further

One's level of computing literacy will determine how one will be able to participate in new kinds of social interactions made possible by new ICT. As the examples above have shown, higher literacy levels raise the transformative power of communication to unprecedented standards. The challenge is in the air: "If computational thinking will be used everywhere, then it will touch everyone directly or indirectly. [...] If [it] is added to the repertoire of thinking abilities, then how and when should people learn this kind of thinking and how and when should we teach it?" (Wing 2008: p. 3720)

One of the most successful responses to-date is AgentSheets, a visual programming environment designed mainly for teaching computational thinking at schools (Repenning and Ioannidou 2004). The *Scalable Game Design* Project, carried out by Alexander Repenning and his group at Colorado University at Boulder has been educating school teachers and students in computational thinking for many years now. In 2010, we brought the project to Brazil and started working with a public school in the city of Niterói.

Given our interest in semiotic theories of HCI, we have approached the project with a dual perspective. On the one hand, we want to educate teachers and students in computational thinking. On the other, we want to show them that computational thinking gives them a completely new *language* for self-expression and communication, with which they can do whatever they can *mean* (or "signify", in semiotic terms). Our strategy is to work constantly with the semiotic relations that hold between three distinct signification systems (see Figure 25.13): natural language, game play (*i. e.* executable games programmed by AgentSheets users) and game code (*i. e.* the various levels of programming language encodings that make the games *run*). In AgentSheets, the users can program in Visual AgenTalk, a visual programming language with a textual XML counterpart.

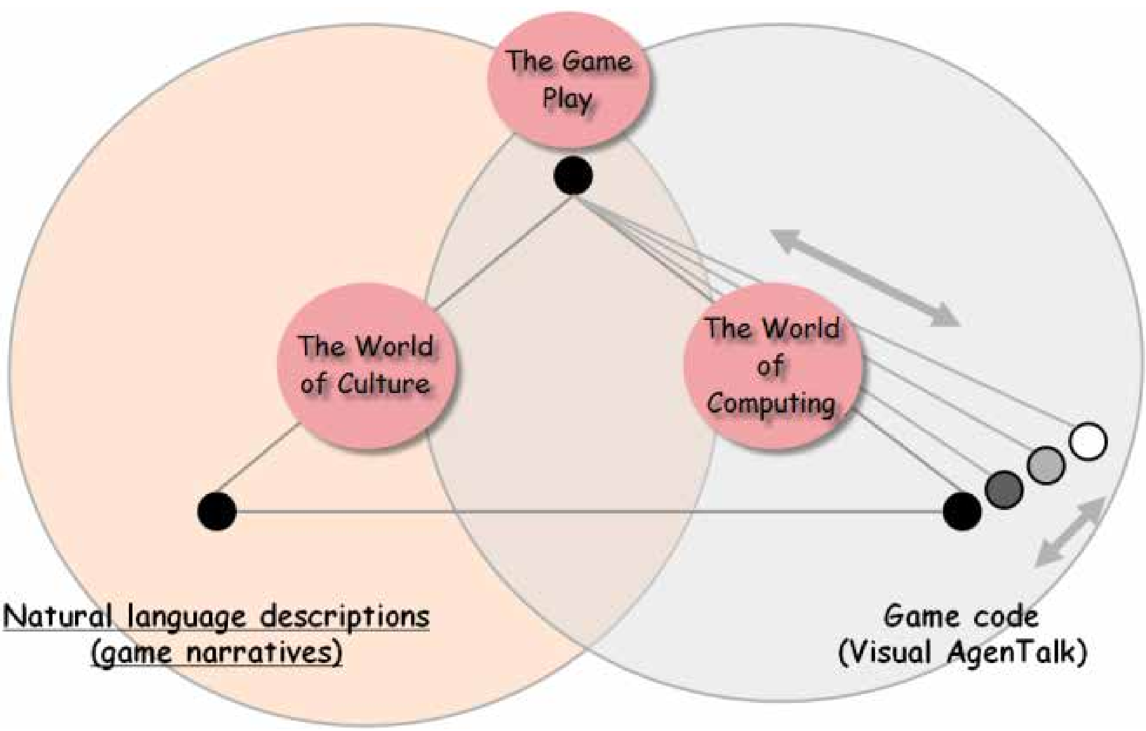



Figure 25.13: Semiotic varieties in a visual programming environment.



Empirical evidence collected in game design sessions with a group of students in 2010 has shown elaborate and intriguing signification patterns when we compared, for each game, the natural language descriptions provided by the *programmers*, the Visual AgenTalk code, and the resulting executable game representations (de Souza et al 2011). In particular, we traced interesting entity-naming strategies, token/type relations in representation choices, and curious transitive structure changes when contrasting natural language and computer program representations. Subsequent research steps carried out in 2011 with another group of students suggest that semiotic relations between signification systems involved in using AgentSheets can be used to raise the level of *computing awareness* among learners.

In Figure 25.14, we show successive screen shot snippets from a semiotic exploration of relations between *game play* signs, *game code* signs and *natural language* signs in a very abstract environment used by CS graduate students. The "game" representation (an executable simulation) has only four static signs: an orange circle; three colored squares; and a black background. The dynamic signs are also very few and very simple: if the user presses arrow keys the orange circles moves in the corresponding direction; if the circle moves next to the boundaries of any of the three squares, it is "trapped" by the square and the game is "reset". The fascinating aspects of the exploration were: to examine what *meanings* the participating students assigned to the simulation; and to examine how the apparently identical behaviors of the three squares had been intentionally encoded by the experimenters as completely different representations.


FIGURE 25.14: Screen shots from a semiotic exploration of simulation signs with AgentSheets.



When asked to write down their explanation ("Explicação" in Figure 25.14) for the orange circle behavior, inferred exclusively from the participant's interaction with the game, the students used signs referenced to the "protagonist" of the game. For example, the explanation depicted in Figure 25.14 says that "the orange circle cannot collide with any of the obstacles represented by the squares". A totally different story, however, would be told if the student looked at the underlying program. The three squares are actually programmed as *very different* agents. The green square "traps" the circle as it comes near it. The white square is like an attractive area into which the orange circle "throws itself" as soon as it comes near it. And the white square is no more than "a hole in the ground", an agent-free space in the worksheet, into which the circle may "fall propelled by the ground". Thus, the *designer's* story about this simulation is a totally different one if looked at from the inside or the outside. From the inside, it might go like this, for instance: "The orange circle loves the grey square and jumps into it as soon as it sees it; it must escape from the green square that traps it when the circle comes near it; and it must beware of a square-shaped hole in the ground, into which it must not fall."

An HCI-oriented mind will probably say that for the designer to *communicate* his story appropriately to the user, the abstract squares should be replaced with *icons* that adequately represent "trap", "attractive area" and "hole in the ground". However, this is actually telling only part of the computational thinking story, of course. The fact that similar *effects* can be computationally represented by very different *data structures and algorithms* is one of the most fundamental and powerful principles of Computer Science. Therefore, we *should* be telling school teachers and students that computational signs are like plastic semiotic material, and stimulating them to play with it in as many ways as they can imagine. This is what will really empower them to *express themselves computationally*.

In Figure 25.15, we show a sketched version of *Exploratorium*, a semiotically-rich prototype plug-in to AgentSheets, which we are developing for its users to be able to explore how different game play representations are *encoded* in Visual AgenTalk. Our intent is to raise the students' and teachers' awareness of the various possibilities in computer representations, and to stimulate them to play with these possibilities at their will. The links on the illustrated page show that users can ask a number of things about the representations *they see* (how they are produced, where they are located, what they *stand for* and why). The answers are partly generated automatically from Visual AgenTalk representations and partly retrieved from annotations (secondary notations) elicited from the programmers in dialogs *about what they meant*.

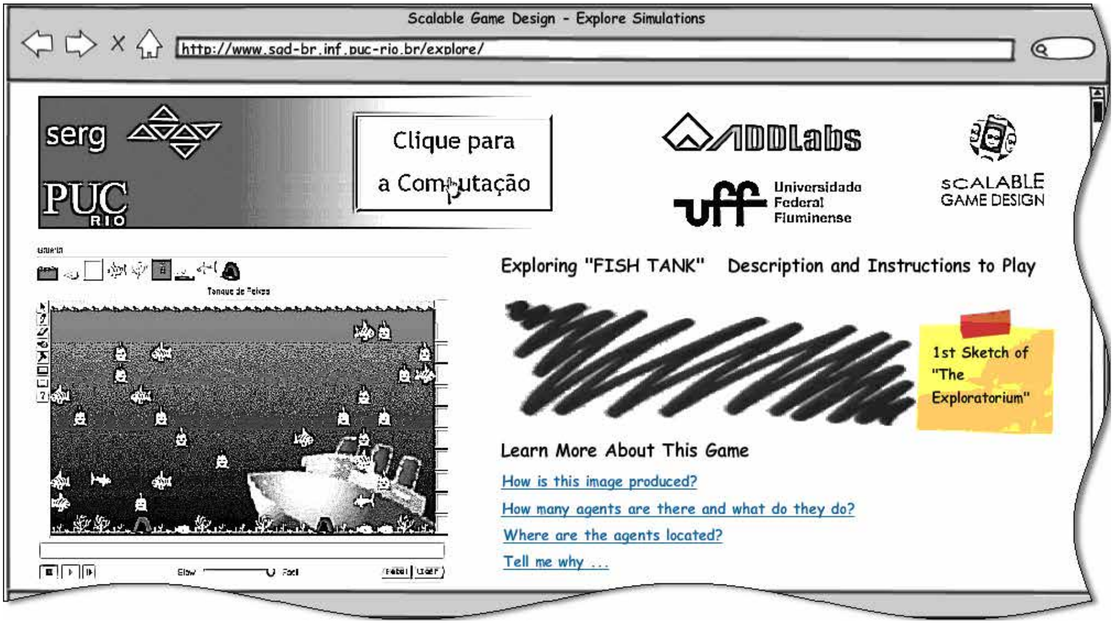



Figure 25.15: The Scalable Game Design Brasil Project: A semiotically-rich environment to explore signs in computing.



In Figure 25.16, we show a sketched version of the *Web Navigation Helper* (Intrator and de Souza 2009; Intrator and de Souza 2008; Monteiro and de Souza 2011), a user agent to increase Web accessibility that we have been expanding to bridge cultural barriers of various sorts (language barriers, literacy barriers, etc.). As shown in this illustration, the user agent actually *rephrases* the designer's communication originally expressed in a Web page. In Figure 25.16, the user agent is an *interpreter* for a Brazilian user that is trying to navigate a Web page produced in the USA. Technically, the user agent is a Firefox extension that works in combination with **CoScripter**. For every interaction that it *mediates*, the agent requires a *recorded CoScript* and a full *specification of the mediating dialog* that will take place at each step of the recorded interaction.

The illustrated dialog, in Portuguese, is telling the Brazilian user how he or she can interact with the *Scalable Game Design Arcade* page. Notice that the dialog is not a *verbatim* translation of the page (although this would *also* be a kind of mediation). The agent is actually *guiding* a foreigner's interaction with material produced in and by another *culture*.

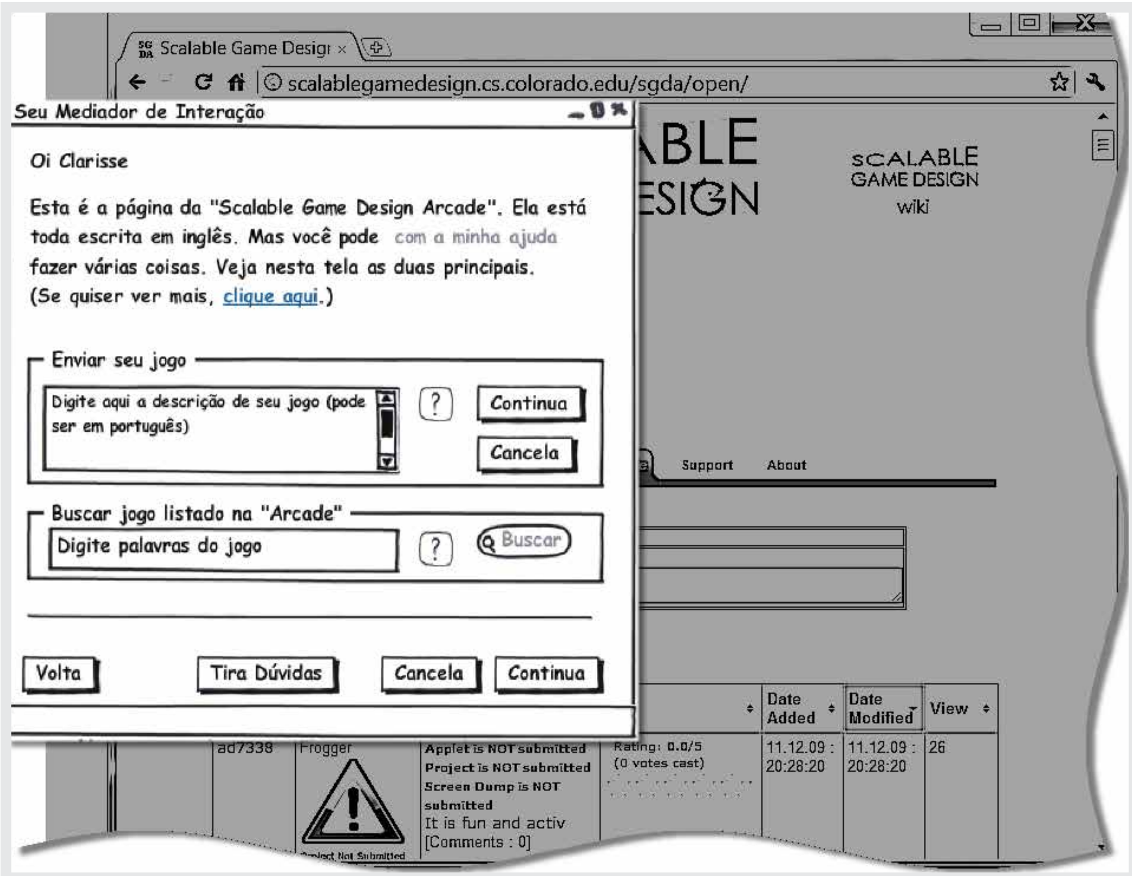



Figure 25.16: An interaction mediator helping a Brazilian user navigate through a website in English.



Guidance is designed using *conceptual viewpoint metaphors* proposed by Salgado (Salgado et al 2011; Salgado et al 2012). Very briefly, the metaphors allow interaction designers to organize cross-cultural communication from different perspectives, ranging from the strongest form of mediation (where the foreign culture is *invisible* to the user thanks to communication carried out in his or her own native language, with complete neutralization of all references to *foreign* cultural practices and values) to the weakest, actually non-existing, form of mediation (where the foreign culture is fully exposed, with *no* trace of the user's culture — communication is carried out in a foreign language, with reference to foreign cultural practices and values).

Semiotic Engineering research, as the examples above can show, expands the limits of user-centered HCI and establishes connections with other specialized areas of Computer Science and Information Systems Development, like Programming (through the semiotics of computer representations) and Information Architecture (through the organization of explicit cultural references to be used in computational signification systems).

## 25.6 SO, WHAT'S IN IT FOR ME?

In his preface to the *Encyclopedia of Semiotics* Paul Bouissac writes:

> "*The twentieth century has witnessed an increasing fragmentation of knowledge into a multitude of disciplines and specialties. At the same time, integrative visions have arisen in an effort to make sense of the flood of information generated by the modern intensification of formal and empirical research. Semiotics represents one of the main attempts — perhaps the most enduring one — at conceiving a transdisciplinary framework through which interfaces can be constructed between distinct domains of inquiry.*"
>
> *-- Bouissac 1998: p. ix*

Only a few years after Bouissac, John Carroll writes in the introduction to one of his books:

> "*An ironic downside of the inclusive multidisciplinarity of HCI is fragmentation. [...] There are too many theories, too many methods, too many application domains, too many systems. Indeed, fragmentation may be a bit worse than it has to be. Some HCI researchers, faced with the huge intellectual scope of concepts and approaches, deliberately insulate themselves from some portion of the field's activity and knowledge. This tension between depth and breadth in scientific expertise is not unique to HCI but it clearly undermines the opportunity for multidisciplinary progress.*"
>
> *-- Card et al 1983: p. 6*

So, what's in Semiotics for you?

I hope that this chapter has provided multiple *signs* of my personal answer to the question. Semiotics can provide the foundations for new integrative theories of HCI. Now, why should we need integrative theories? One of the answers is explicitly given by John Carroll in the citation above. But this chapter has touched on some of the implications of Carroll's answers, as well as on some other possible answers to the challenging question.

A semiotic perspective on HCI proposes that *meaning* is the most fundamental concept in this discipline, and that our inquiry should account for how meaning is *communicated* through computer systems interfaces, in as many guises as they now occur. It follows from this that HCI must not only care for *meaning takers*, but also for *meaning makers*: communication implies an *exchange* of signs produced by at least two communicating *minds*. Therefore, to escape the embarrassment of having to postulate that *systems* have minds of their own, HCI will benefit from theories that can account for meaning-making processes in which HCI designers clearly engage while in the process of building systems interfaces to support productive user interaction with computing artifacts. In fact, computer meaning-making process can engage not only designers, but also programmers and a whole host of developers that encode meaning elements (signs) in software tools with which modern computer artifacts are built. Therefore, a semiotic perspective can potentially *track* human signification throughout long chains of software instances, in space and time, just like semiotic theories trace meaning in society, culture and history.

Because they push the designers into representing *self*-images and thinking of their communicative *intent and strategies*, HCI theories of semiotic extraction can also stimulate the development of a new *ethics* in HCI design. Although the UCD tradition has clearly established an ethics *commitment* with the users' needs and aspirations, it has not contributed much to an ethics of *communication* since UCD theories only talk about relations and interactions between users and *systems* — not about users and systems *designers*. It is remarkable, however, that in his most recent book (Norman 2010), Don Norman, who inspired UCD with his Cognitive Engineering and seven-step theory of action in the 1980's (Norman 1986) speaks openly of the importance of communication to live with complexity, and goes as far as dedicating a full chapter to "social signifiers" — a semiotic concept. This is itself the sign of that an *ethics of communication* begins to make its way into design and HCI.

This new ethics may play a relevant role in the wake of the 21st century when global awareness of the need to respect and preserve cultural diversity is greater than ever before. A semiotic perspective can help technologists in trying to prevent that their products inadvertently extinguish valued diversity of meanings from user communities (Salgado et al 2012).

By bringing designers and users together in the process of HCI, a semiotic perspective can naturally promote more *reflection in action* (Schön 1983). When designers view themselves as *talking to users*, they can naturally tap on their instinctive abilities to *signify* reality and prompt communicative reaction from people around them. They are as *native* in the "human" side of *human-computer interaction* as the users. Therefore, we need as many good theories about *them* as we have about *users*. *Reflective* theories will help designers to produce computer-mediated communication in which they can consciously model *their* own signification and communication

A semiotic outlook can also bring about some clarifications at *meta-theoretical* levels that scientific research should attend to. In this chapter, we have only highlighted aspects of Cartesian influence in the thinking of many HCI researchers and practitioners. We commented on why ultimate *primary* cognitions, which human minds can access by means of a rigorous inspection method, are not only appealing to computer language theorists, but often tacitly subscribed by members of the HCI community. This notion allows for framing HCI as a formal linguistic exchange between humans and computers. Semiotic theories that are philosophically aligned with this specific Cartesian view will probably reinforce beliefs that there are primitive conceptual meanings that can be known in the process of investigation.

However, even semiotic theories that do not 'upset the applecart' in CS, as non-Cartesian theories tend to do, lead researchers into making explicit commitments with the ontological origins of meaning and its variations. For example, a semiotic foundation requires that researchers provide a theoretical explanation of why and how people (in our case *users* and *designers*) assign different meanings to identical representational stimuli. This explanation will guide their choices in designing computer-mediated communication between software producers and software users, showing that theoretically-informed design is actually a lot more than *a craft*. Researchers will also have to deal with hard methodological and epistemological questions. If there are such things as conceptual meaning primitives, what are they? How can they tell *true* primitives from *false* ones? How can they explain that sometimes our meanings are *right*, and other times these very meanings are *wrong*? How can they explain why, even when some of their meanings are wrong, people occasionally make the *right* inferences using them? All of these questions are fundamentally important for scientific research in HCI and, as we have seen, Semiotics can play an important role in helping us to formulate them, to understand what *they mean*, and then to investigate possible answers.

To conclude, it is my belief that, grounded on more explicit philosophical commitments, the most important contribution of Semiotics to HCI is to rehabilitate human talent and intuition from the dark corners where it has been hiding, and to bring it to light as one of the most powerful elements for effective, efficient, creative and enjoyable computer-mediated communication in the 21$^{st}$ century. Our task, as HCI researchers and practitioners, is to generate and use the knowledge that is necessary to turn computers into the richest *medium* for human signification and communication.

## 25.7 ACKNOWLEDGMENTS

I have to thank my sponsors — CNPq and FAPERJ — and many people that have supported, stimulated and developed research in Semiotic Engineering over the last twenty years. In particular, I would like to thank namely those that have contributed more directly to SERG projects mentioned in this chapter: Carla Leitão, Luciana Salgado, Cristina Garcia, Ingrid Monteiro, Juliana Ferreira, Marcelle Mota, Cleyton Slaviero, Eduardo Tolmasquim, Leonardo Faria, Alexander Repenning, Nadia Repenning, Allen Cypher, Gerhard Fischer, CoScripter developers and AgentSheets developers. I would also like to thank Mads Soegaard and Allan Holstein-Rathlou for first-class editorial work. I owe special thanks to Mads for having invited me to write this chapter and for his constant encouragement. It has

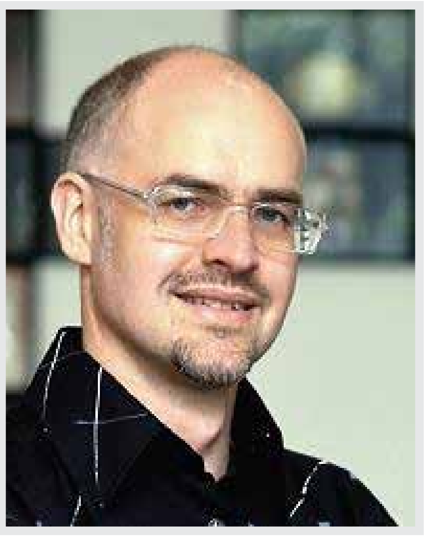
been a much longer project than I first thought it would be, but also a much more rewarding thinking and learning adventure.

For more information about Semiotic Engineering, please visit our website.

## 25.8 COMMENTARY BY ALAN BLACKWELL

How to cite this commentary in your report

### Alan Blackwell



*© Alan Blackwell*

I only have one big research question, but I attack it from a lot of different angles. The question is representation. How do people make, see and use things that carry meaning? The angles from which I attack my question include various ways in which representations are applied (including design processes, interacting with technology, computer programming, visualisation), vario...

Alan Blackwell
Alan Blackwell is a member of The Interaction Design Foundation

Clarisse Sieckenius de Souza provides a superbly lucid introduction to the fundamental concepts of semiotics, and their relevance to interaction design, just as we would hope from the author of her influential textbooks introducing Semiotic Engineering. These are principles that are directly relevant to students and professional software designers, principles that many computing specialists find intriguing but also challenging. However her chapter carries a far more profound challenge to readers – a challenge that was already implicit when she juxtaposed the world views of semiotics and of engineering in her book titles. Where engineering methods have traditionally relied on materials whose properties and behavior can be predicted from immutable laws of physics or chemistry, the “materials” of interactive software are socially – semiotically – constructed, and always changing. When engineers are forced to confront computers as media, many engineering fundamentals appear to be threatened at the same time. This tension is likely to remain a feature of HCI research and practice for as long as HCI remains a branch of Computer Science.

But an even larger challenge looms in the way that we all choose to respond in future to the pervasiveness of computing technology. The campaign for Computational Thinking responds, quite rightly, to the demand for educated citizens who can engage in dialogue around computational media, rather than being passive consumers (if any such thing were ever possible in a semiotic world). One consequence of that dialogue is to reconceptualise the notion of “programming”, such that it becomes more than a simple engineering technique, but rather a mode of participation in society. This is a warning for those tempted to assume that computational thinking is simply the evangelical wing in the church of computer science. As shown by the pioneering work in end-user programming that de Souza presents from Cypher, Repenning and others, the ability to configure and control software assists all technology users to recognize the semiotic structures of the society we are building.

## 25.9 REFERENCES


(ed.) (1998): *The Essential Peirce, Volume 2: Selected Philosophical Writings, 1893-1913.* Indiana University Press

Andersen, Peter B. (1997): *A Theory of Computer Semiotics : Semiotic Approaches to Construction and Assessment of Computer Systems.* Cambridge University Press

Andersen, Peter B. (1990): *A Theory of Computer Semiotics.* Cambridge, UK, Cambridge University Press

Andersen, Peter Bøgh, Holmqvist, Berit and Jensen, Jens F. (eds.) (1994): *The Computer as Medium (Learning in Doing: Social, Cognitive and Computational Perspectives).* Cambridge University Press

Andersen, Peter Bøgh, Holmqvist, Berit and Jensen, Jens F. (eds.) (1993): *The Computer as Medium (Learning in Doing: Social, Cognitive and Computational Perspectives).* Cambridge University Press

Baranauskas, M. Cecilia M., Liu, Kecheng and Chong, Samuel (2003): Website Interfaces as Representamina of Organizational Behaviour. In: Gazendam, H. W., Jorna, René and Cijsouw, R. S. (eds.). "Dynamics and Change in Organizations: Studies in Organizational Semiotics (Studies in Organisational Semiotics, 3.)". Springer

Beyer, Hugh R. and Holtzblatt, Karen (1998): *Contextual Design: Defining Customer-Centered Systems.* San Francisco, Morgan Kaufmann Publishers

Bonacin, Rodrigo, Baranauskas, M. Cecili C. and Liu, Kecheng (2004): Interface Design for the Changing Organisation: an Organisational Semiotics Approach. In: Liu, Kecheng (ed.). "Virtual, Distributed and Flexible Organisations: Studies in Organisational Semiotics". Springer

Bouissac, Paul (ed.) (1998): *Encyclopedia of Semiotics.* Oxford University Press, USA

Buchler, Justus (ed.) (1955): *Philosophical Writings of Peirce.* Dover Publications

Burrell, Gibson and Morgan, Gareth (1979): *Sociological Paradigms and Organisational Analysis: Elements of the Sociology of Corporate Life.* Ashgate Publishing

Burrell, Gibson and Morgan, Gareth (2007): *Sociological Paradigms and Organisational Analysis: Elements of the Sociology of Corporate Life.* Ashgate Publishing

Card, Stuart K., Moran, Thomas P. and Newell, Allen (1983): *The Psychology of Human-Computer Interaction.* Hillsdale, NJ, Lawrence Erlbaum Associates

Carroll, John M. (ed.) (2003): HCI Models, Theories, and Frameworks: Toward a multidisciplinary science. San Francisco, Morgan Kaufmann Publishers

Carroll, John M. and Rosson, Mary Beth (1987): The paradox of the active user. In: Carroll, John M. (ed.). "Interfacing Thought: Cognitive Aspects of Human-Computer Interaction (Bradford Books)". The MIT Presspp. 80-111

Cypher, Allen (1993): Watch What I Do: Programming by Demonstration. Cambridge, MA, The MIT Press

de Souza, Clarisse Sieckenius (1993): The Semiotic Engineering of User Interface Languages. In International Journal of Man-Machine Studies, 39 (5) pp. 753-773

de Souza, Clarisse Sieckenius and Cypher, Allen (2008): Semiotic engineering in practice: redesigning the CoScripter interface. In: Levialdi, Stefano (ed.) AVI 2008 - Proceedings of the working conference on Advanced Visual Interfaces May 28-30, 2008, Napoli, Italy. pp. 165-172

de Souza, Clarisse Sieckenius and Leitäo, Carla Faria (eds.) (2009): Semiotic Engineering Methods for Scientific Research in HCI (Synthesis Lectures on Human-Centered Informatics). Morgan and Claypool Publishers

Descartes, Rene (2004): Discours de la Methode. Project Gutenberg

Eco, Umberto (1978): A Theory of Semiotics. Indiana University Press

Eco, Umberto (1976): A Theory of Semiotics. Indiana University Press

Fischer, Gerhard (2011): Understanding, fostering, and supporting cultures of participation. In Interactions, 18 (3) pp. 42-53

Floch, Jean-Marie (ed.) (2001): Semiotics, Marketing and Communication: Beneath the Signs, the Strategies (International Marketing Series). Palgrave Macmillan

Fogg, B. J. (2002): Persuasive Technology: Using Computers to Change What We Think and Do. Morgan Kaufmann

Hirschheim, Rudy, Klein, Heinz K. and Lyytinen, Kalle (1995): Information Systems Development and Data Modeling: Conceptual and Philosophical Foundations. Cambridge University Press

Hix, Deborah (ed.) (1988): Advances in Human-Computer Interaction. Norwood, NJ, Intellect

Hjelmslev, Louis (1961): Prolegomena to a Theory of Language. Univ of Wisconsin Press

Hopcroft, John E., Motwani, Rajeev and Ullman, Jeffrey D. (1979): Introduction to Automata Theory, Languages, and Computation. Addison Wesley

Hopcroft, John E., Motwani, Rajeev and Ullman, Jeffrey D. (2006): Introduction to Automata Theory, Languages, and Computation (3rd Edition). Addison Wesley

Hopcroft, John E., Motwani, Rajeev and Ullman, Jeffrey D. (2000): Introduction to Automata Theory, Languages, and Computation (2nd Edition). Addison Wesley

Houser, Nathan and Kloesel, Christian J.W. (eds.) (1992): The Essential Peirce, Volume 1: Selected Philosophical Writingsâ (1867-1893). Indiana University Press

Intrator, Chantal and de Souza, Clarisse Sieckenius (2008): Using web scripts to improve accessibility. In: Proceedings of the VIII Brazilian Symposium on Human Factors in Computing Systems 2008. pp. 292-295

Kammersgaard, John (1988): Four Different Perspectives on Human-Computer Interaction. In International Journal of Man-Machine Studies, 28 (4) pp. 343-362

Kaptelinin, Victor and Nardi, Bonnie A. (2006): Acting with Technology: Activity Theory and Interaction Design. The MIT Press

Khaled, Rilla, Barr, Pippin, Fischer, Ronald, Noble, James and Biddle, Robert (2006): Factoring culture into the design of a persuasive game. In: Kjeldskov, Jesper and Paay, Jane (eds.) Proceedings of OZCHI06, the CHISIG Annual Conference on Human-Computer Interaction 2006. pp. 213-220

Light, Ann (2001): Interactivity and User Commitment - Relationship Building through Interaction on Websites. In: Proceedings of the HCI01 Conference on People and Computers XV 2001. pp. 459-474

Liu, Kecheng (2000): Semiotics in Information Systems Engineering. Cambridge, Cambridge University Press

Monteiro, Ingrid Teixeira and de Souza, Clarisse Sieckenius (2011): Embedded cultural features in the design of an accessibility agent for the web. In: Proceedings of the 6th international conference on Universal access in human-computer interaction design for all and eInclusion - Volume Part I 2011. pp. 295-304

Morris, Charles (1971): Writings on the General Theory of Signs. Mouton De Gruyter

Mullet, Kevin and Sano, Darrell (1995): Designing Visual Interfaces: Communication Oriented Techniques. Sunsoft Press

*Nadin, Mihai (1988a): Interface design: A semiotic paradigm. In Semiotica, 69 (3) pp. 269-302*

*Nadin, Mihai (1988b): Interface design and evaluation - Semiotic implications. In: Hix, Deborah (ed.). "Advances in Human-Computer Interaction". Norwood, NJ: Intellectpp. 45-100*

*Norman, Donald A. (2009): Systems thinking: a product is more than the product. In Interactions, 16 (5) pp. 52-54*

*Norman, Donald A. (1986): Cognitive engineering. In: Norman, Donald A. and Draper, Stephen W. (eds.). "User Centered System Design: New Perspectives on Human-Computer Interaction". Hillsdale, NJ: Lawrence Erlbaum Associatespp. 31--61*

*Norman, Donald A. (2010): Living with Complexity. The MIT Press*

*Norman, Donald A. (2007): The Design of Future Things. Basic Books*

*Norman, Donald A. (2004). Design as communication. Retrieved 19 March 2012 from JND.org: http://jnd.org/dn.mss/design_as_communication.html*

*Norman, Donald A. and Draper, Stephen W. (eds.) (1986): User Centered System Design: New Perspectives on Human-Computer Interaction. Hillsdale, NJ, Lawrence Erlbaum Associates*

*Néris, Vânia Paula de Almeida, Almeida, Leonelo Dell Anhol, Miranda, Leonardo Cunha de, Hayashi, Elaine Cristina Saito and Baranauskas, M. Cecília C. (2011): Collective Construction of Meaning and System for an Inclusive Social Network. In International Journal of Information Systems and Social Change (IJISSC), 2 (3) pp. 16-35*

*O'Neill, Shaleph (2008): Interactive Media: The Semiotics of Embodied Interaction. Springer*

*Prinsloo, Mastin and Walton, Marion (2008): Situated Responses to the Digital Literacies of Electronic Communication in Marginal School Settings. In Yearbook 2008 African Media African Children, pp. 99-116*

*Repenning, Alexander and Ioannidou, Andri (2004): Agent-based end-user development. In Communications of the ACM, 47 (9) pp. 43-46*

*Saussure, Ferdinand de (1972): Cours de linguistique générale. Payot*

*Schön, Donald A. (1988): Educating the Reflective Practitioner: Toward A New Design for Teaching and Learning in the Professions. San Francisco, Jossey-Bass*

Schön, Donald A. (1983): The Reflective Practitioner: How Professionals Think In Action. Basic Books

Souza, Clarisse Sieckenius de de (2005): The Semiotic Engineering of Human-Computer Interaction (Acting with Technology). The MIT Press

Spool, Jared M. and Schaffer, Eric M. (2005): The great debate: can usability scale up?. In: Proceedings of ACM CHI 2005 Conference on Human Factors in Computing Systems 2005. pp. 1174-1175

Umiker-Sebeok, Jean (ed.) (1988): Marketing and Semiotics: New Directions in the Study of Signs for Sale. Mouton De Gruyter

Wing, Jeannette M (2008): Computational thinking and thinking about computing. In Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 366 (1881) pp. 3717-3725

Winograd, Terry (1996): Bringing Design to Software. ACM Press

Winograd, Terry and Flores, Fernando (1986): Understanding Computers and Cognition. Norwood, NJ, Intellect

Zemanek, Heinz (1966): Semiotics and programming languages. In Communications of the ACM, 9 (3) pp. 139-143

## About the author

### Clarisse Sieckenius de Souza

*© Clarisse Sieckenius de Souza*

Has also published under the name of:
"Clarisse de Souza"

Personal Homepage:
www-di.inf.puc-rio.br/~clarisse/

Clarisse Sieckenius de Souza is Full Professor of Computer Science at the Departamento de Informática, PUC-Rio, where she has also served as the director of Graduate Studies Program for two terms (2003-2005 and 2007-2009). In 2010, Clarisse was a co-winner of the prestigious ACM SIGDOC Rigo Award (see pictures and more in SERG Website). In 1996 she founded SERG (the Semiotic Engineering Research Group). Among the 15 M.Sc. and 19 Ph.D. students that she has (co-)supervised, more than a half are faculty in various Brazilian universities, and nearly one third work in the industry. In 2005 she published her first book, The Semiotic Engineering of Human-Computer Interaction (The MIT Press). Her second book, Semiotic Engineering Methods for Scientific Research in HCI, co-authored by Carla Leitão, was published in 2009, in Morgan & Claypool's Synthesis Lectures Series.

## Your notes and thoughts on chapter 25

Record your notes and thoughts on this chapter. If you want to share these thoughts with others online, go to the bottom of the page at:
http://www.interaction-design.org/encyclopedia/semiotics_and_human-computer_interaction.html

### NOTES

# Chapter 26

# Aesthetic Computing

*by Paul A. Fishwick.*

The phrase “Aesthetic Computing” while taken literally applies the philosophical area of aesthetics to the field of computing, and work in the area is broadly defined as such; however, in my operational definition for the work we do in my research lab and in teaching, aesthetic computing is treated as *embodied formal language*. The purpose of aesthetic computing is to deliver knowledge and practice of formal languages using aesthetic products as a vehicle. Aesthetic Computing is founded on an increasing collection of literature on the role of the body in learning, specifically in mathematics. This foundation is then applied to the field of computing whose formal language elements are extensions of mathematics. There are two questions that this new area raises:

- Q1: *How can embodied cognition be situated within formal languages?*

- Q2: *How can embodied cognition result in novel computer interfaces for formal languages?*

Q1 surfaces a host of sub-questions revolving around theory, philosophy, and analysis. Asking this question raises issues of motivation: 1) Why am I interested in this topic? 2) How is the area of aesthetic computing built on top of embodied cognition and philosophy? 3) Who has worked in this area (e.g., the literature)? Q1 is not enough, however. It is one matter to analyze and develop theory, but another to ask oneself, "How can this theory be transformed into practice?" That is the essence of Q2. What should we be *doing, practicing*, *and creating* to take embodied cognition of mathematics and computing to the next level? We need to build a new generation of human-computer interfaces that are informed by embodied principles and use these principles as design elements for interacting with formal languages. A potential, and vital, third question would revolve around the effects on such computer interfaces on learning via assessment and scientifically-based research methods. This represents an area that aesthetic computing needs to investigate; however, most work to date is based on theory construction and engineering the novel interfaces.

"*The Aesthetic Computing Hypothesis is that given the embodied nature of cognition, we should realize this embodiment through novel human-computer interfaces for learning formal languages.*"

## 26.1 CONTEXT FOR AREA OF AESTHETIC COMPUTING

I pose two questions as a means to provide context for the area of aesthetic computing: 1) Why is the term "Aesthetic Computing" being treated as "Embodied Formal Languages?" and 2) What are "Embodied Formal Languages"? For the first question, we must revisit the roots of the word "aesthetics." The original Greek definition of aesthetics, *αἰσθητικός* (*aisthetikos)*, stems from another Greek word aisthanomai, meaning "I perceive, feel, sense." At the core of aesthetics, then, lies the body, and its interactions in forming concepts and knowledge: a esthetics as embodiment. Aesthetics is, in breadth and depth, a much richer enterprise above this level (Kelly 1998), yet we maintain a view of aesthetics that is body-based, even though Diffey (1995) notes that the term 'aesthetic' has largely lost its perceptual sense except in the word 'anaesthetic,' but retains its senses of "beautiful' and 'artistic.'" As far as to why "Formal Languages" are used to characterize "Computing," we note that the bulk of theory of automata and computing is situated within linguistics – although a subset of general linguistics that requires a formally well-defined specification and treatment.

Let us now consider the definitions of embodiment and formal language. Embodiment suggests the perception/action feedback loop present when the body interacts with its environment. So, it seems clear that an embodied approach to anything would involve sensorimotor functions – using the mouse, keyboard, multi-touch displays as well as donning a head-mounted display or using a tactile feedback device. Human-Computer Interaction is chock-full of approaches that leverage such technologies. But, embodiment is a much deeper concept than sensory stimuli and physical manipulation. We have a sense of presence with certain advanced technologies such as multi-user virtual environments (i.e., achieving different types of presence, including social). We also have a sense of presence when reading a book since the book situates our "mind's body" within the narrative (ref. "narrative psychology" in Beck et al. 2011). Thus, embodiment can be measured objectively by hardware used to enable the senses, or subjectively through a presence instrument on the human subject. Embodiment should not be viewed as a rejection of abstraction, but rather as a complement to it (Devlin 2006).

Formal languages define a category of language that is artificial, such as a programming language. These languages stem from formal grammars which can be based on text, shapes, or diagrams. FORTRAN, Java, and Perl are examples of formal languages, but so are the eXtensible Markup Language (XML), Unified Modeling Language (UML), data structures, Morse code, and dynamic model structures used for simulation (Fishwick 1995, Fishwick 2007b). Formal languages are frequently specified using grammars such as the Backus-Naur Form (BNF) and need not be text-based. For example, one can have formal audiovisual languages and also graph grammars. All formal language structures can be defined hierarchically using levels of abstraction (e.g., 3 finite state machine levels governing an underlying set of ordinary differential equations, which in turn are translated into the programming language Java, and then further into byte code). Languages, therefore, are frequently defined in long chains of specification and translation. Each language has its own target functionality, culture, and adherents. Ghezzi and Jazayeri (1997) provide general concepts of specification for programming languages.

## 26.2 PERSONAL EXPERIENCES AND INFLUENCES

### 26.2.1 Art

It is easy to take the idea of embodied cognition for granted since it seems like something so natural—that the body plays a central role in cognition. However, an adherence to embodiment tends to change your worldview when looking at objects. As an amateur artist, I collected many posters and prints of historically well-known artists. In middle school, I was strongly influenced by Thomas Gainsborough's work, in particular Figure 26.1.

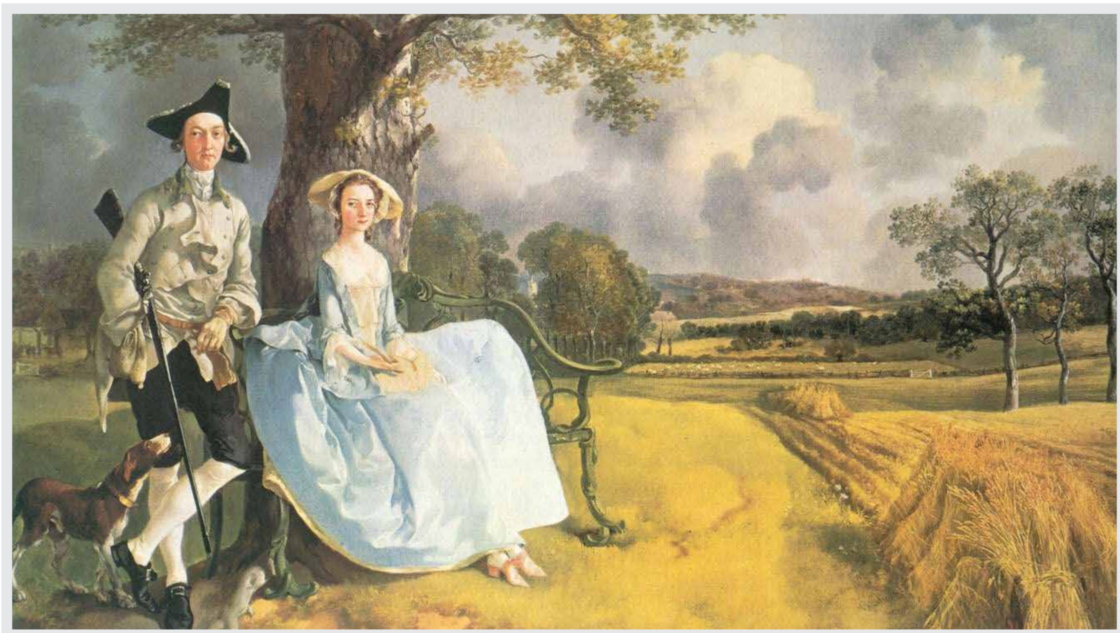


Figure 26.1: Mr. and Mrs. Andrews, oil on canvas, Thomas Gainsborough, 1750. The National Gallery, London, UK.



I imagined that with myself as an avatar , I could enter the painting, walk the wheat field, examine the trees, and engage in social discourse with Mr. and Mrs. Andrews. This led to a series of imaginary conversations and observations "in world." The key point here is the "reading" of this work *as a form of embodied experience*. The Gainsborough painting was not a remote object of study for me, but rather an example of virtual reality, a time machine—an illusion that allowed me to immerse myself within the world of 18th century England. This approach is an example of Dewey's *art as experience* (Dewey 1934) and relates to Grau's (Grau 2004) argument about artists as the first virtual reality creators. The approach stresses that when we approach an object, we can interpret it dynamically via a bodily simulation with all of the perceptual and motor-based actions that the body affords. This way of thinking and acting can be applied to all objects and media, including mathematics and computing.

### 26.2.2 Mathematics

In elementary school, like hoards of other students throughout the world, I was taught the elements of arithmetic – its methods and laws, with many examples that were exercised using rote memorization and intense practice. Doing mathematics was highly action-based, but the action was limited to solving multiple problems over extended periods of time. After the basic elements of arithmetic came algebra. Let's consider the following mathematical expression containing arithmetic with a sliver of algebra:

"*X = 2 * (3 + 4).*"

We have all been subject to such mathematical objects as they are critical to an educated public. Learning all components of this equation was not easy – one had to understand the concept of a variable, operations of multiplication and addition, followed by the concept of a parenthetically-delimited group. Order of operations is also critical, as suggested by the group. So, for example, I can add 3 to 4 and then multiply by 2 to obtain 14, which was then set to X as an equivalence. Certain laws of arithmetic were useful in transforming expressions such as this one. The Law of Distribution states that `x(y + z) = xy + xz` where `x`, `y`, and `z` are numbers, and the multiplication is implicit rather than being defined explicitly using `*` as in the above equation. The teacher would define the law of distribution and give us many useful examples as a means to reinforce our understanding of the law and how it can be employed in symbolic manipulation. Such patterns of equivalence drove a static pattern-matching type of approach to mathematics.

However, during the ensuing lessons, I found it convenient to create an artificial method of solution that involved treating the numbers and symbols as physical objects. In mathematics education, this kind of process is termed reification (Sfard 1994) and is related to constructivism (Piaget 1950) and constructionism (Papert 1980), where students create their own knowledge through a combination of ideas and life experiences. I used a virtual manipulation of the above expression by representing the distributive law through analogy and metaphor:

"Grab the "2" object, which when juxtaposed with the "*" operator, provides a biomechanical state where the "2" is pushed inward toward the group object defined within the parentheses "(...)". The "2" is pushed gradually and then when it reaches the edge of the spatial boundary denoted by "(", it moves through it to the other side and splits – in a biological fashion – into two clones that are attached to the "3" and to the "4," respectively. This cloning activity results in the expression `(2 * 3 + 2 * 4)`. The sub-expressions `2 * 3` and `2 * 4` are evaluated through further bodily activity. Pushing the `2` and `3` into the `*`, for example, results in multiplications. Similar reactions occur to perform the `+` operation last, as dictated by the learned order of operations. The result is then placed manually in a box with an `X` printed on it."

Mathematics then, for me, had become akin to a full-body sport rather than simple operations requiring a collection of static text-based rules and patterns. The virtual manipulations might involve other embodied activities, where I might have "launched the " over a wall that bounds the parenthetical expression. While this is a personal experience, it is by no means unique, as Sfard observes in her dialogue with Thompson (Thompson and Sfard 1994), where she notes the propensity for similar mental imagery: "My work with mathematicians brought lots of further evidence that, indeed, the inner world of a mathematizing person may look very much like a material, populated with objects which wait to be combined together, decomposed, moved and tossed around." Arzarello (2004) explains the difference between natural versus formal mathematical presentations, and surfaces the importance of gesture in using naturalistic explanations and interpretations in addition, or on the path, to the formal. The previous embodied description would be termed natural. Goldin and Kaput (1996) overview the effects of media on mathematical representation by noting "..changes in physical media that permit external representations to be action rather than display representations give these representations one characteristic of powerful internal representations." Hadamard (1996) studied mathematical thought which echoed similar cognitive processing. This action-based narrative on mathematical symbols was not limited to the distributive law for me. For example in an expression such as , something interesting happens when moving numbers through the equals sign. There is a virtual line or plane that intersects at a right angle to the . When a number such as is dragged through this vertical plane, the number flips its sign on the other side with a mirror-like effect, resulting in . The laws of commutativity and associativity have similar pseudo-physical, material, behaviors that can be used to understand and process arithmetic expressions.

The problem with my early experiences with embodied sense of symbol manipulation is that none of the books (or teachers) explained mathematics in this way, and I, and likely many others, were forced to keep these somewhat peculiar cinematic episodes to ourselves. Whether this type of thinking is common requires more scientific studies and reflection upon the nature of mathematics. At the University of Florida, we have developed a web-based interactive tool that allows anyone to manipulate expressions in this fashion. We have also previously explored similar embodied representations involving a sense of presence in a virtual environment (Fishwick and Park 2008a).

My purpose of relaying this experience is to emphasize the importance of the body in understanding formal languages such as mathematics. Lakoff and Nunez (2001) presented a landmark compilation of mathematical metaphors that build on top of the philosophy of embodied cognition (Johnson 1987, Varela et al. 1991, Barsalou 2010). In particular, Johnson's image schemata such as containment, attraction, and equilibrium were integral aspects of my arithmetic experience. The literature in embodied thinking centers thought and knowledge on the body and is informed not only by areas such as conceptual metaphor (Lakoff and Johnson 1999, Lakoff and Johnson 2003), but also by subsequent empirical studies of the brain (Feldman and Narayanan 2004, Feldman 2006). Even more generally, language-based narratives appear to contain an embodied basis (Speer et al. 2007, Mar and Oatley 2008) defining natural language in terms of simulation. Reading a story about grasping or running can result in a cognitive simulation of these events and activities, as if the reader had been physically active. Going back in time to when the Method of Loci flourished (Yates 1966), we note that the act of memorizing a set of facts was turned into a rich, embodied process rather than viewed as mere associative retrieval. The area of situated learning and cognition (Brown et al. 1989, Lave and Wenger 1991) meshes well with the embodied approach in terms of its goals and methods: learning by doing.

In closing the discussion of an embodied mathematics, we should note that the concepts of “action”, “interaction”, and “process” can be framed within standard mathematical notation containing explicit aspects of functional composition, dynamics, and procedure (i.e., embodied-types of thought). For example, the aesthetics of geometry and shape can be constructed generatively (Leyton 2001, Leyton 2006) and dynamically via Blum’s wave propagation-based medial axis (Leymarie 2006). We can also use mathematics to create a formal representation of mathematical metaphors (Guhe et al. 2009), thus making a loop: grounding metaphors on mathematical expressions, where the metaphors themselves are formally defined.

The embodied approach has profound implications for mathematics, and by extension for applied mathematics, and computing since computing is a direct outgrowth of mathematics, and formulas such as the one described earlier are common objects found in software “expressions.” If our thought is embodied, then:

- We should investigate the variety of metaphors used within mathematics and computing, and also their origins and cultural associations.
- We should leverage the metaphorical, and embodied, substrate of language by creating new human-computer interfaces that reinforce and amplify this experience.
- We should bring to bear other disciplines for whom “the body” is a natural component, such as the arts and humanities (Slingerland 2008), thus forming new interdisciplinary collaborations that span the academy.

### 26.2.3 Programming

The embodied approach was extended from mathematics into learning programming and data structures. Programming, in particular, is known to be rich in metaphor. Loops are just that: patterns of cyclic behavior – small objects moving around a closed path as these objects perform other tasks. Sequential behavior is sometimes a movement along a spatial path, and functions are machines that take product inputs and produce outputs. Papert (1980) in his explanation of the LOGO language reinforces the importance of embodiment in a term he calls “syntonicity“, where he notes “We have stressed the fact that using the Turtle as metaphorical carrier for the idea of angle connects it firmly to body geometry.” Petre and Blackwell (1999) performed studies on programmers, and results indicate metaphorical reasoning involving objects, motion and general embodied interaction. Metaphors such as these are not only present in all programming languages, but also in the theory of computation on which the theory of computing is based. For example, the Turing machine is an excellent example: a machine envisioned by Alan Turing in the 1930s consisting of a tape read/write head and an infinite tape. This metaphor may have been because of the extensive use of magnetic tape at the time. In the previous century, Charles Babbage used a “mill” in his computing engine. Interestingly, in the vast history of computing where these historical concepts are discussed (Ifrah 2002), most programming and computing was analog and embodied by definition and implementation. It is only relatively recently that the evolution from analog to digital has simultaneously sped up our computations, facilitated a computer revolution, but also disembodied our relations to computing.

### 26.2.4 Media

Media theorists have provided a host of approaches in understanding the evolution of media. McLuhan (1964) places importance, not only on the message created through a modulated medium, but on the medium itself which affects the message. McLuhan employs the example of a light bulb which he claims is a “medium without a message.” However, the light bulb can host a binary digit, and perhaps more in the case of multi-way switch bulbs in a means not unlike Morse code manipulated through signal lamps. Bolter and Grusin (2000) present a theory of media forms undergoing gradual alteration, generally technology-driven, causing us to examine issues of immediacy (seeing beyond the medium to the target signified) and hypermediacy (being aware and reflecting on the medium). New media studies place specific importance on materiality, the medium, and embodiment. Manovich (2002, p. 317), when he considers the “loop as a narrative engine,” with a loop being defined as a common programming structure enabling index-based iteration, asks “Can the loop be a new narrative form appropriate for the computer age?”

Popular media have significantly shaped my thought process underlying aesthetic computing. For example, Tron (Kallay 2011), which debuted in 1982, is noteworthy because it was created based on a highly innovative screenplay which included a large piece of software, namely an “operating system,” that could be experienced directly. Programs were bodies, and the operating system was composed of a city-like space with lighted, moving vehicles and interacting programs.

Tron is fairly unique in this way within the science fiction/fantasy genre. Other more recent cinematic offerings, while impressive and engaging, tend to ignore the "program." For example, on Star Trek: The Next Generation, we were introduced to the Holodeck where one could experience an ultimate virtual reality with full sensory simulation. A user would stop at the outside of the Holodeck and say "Computer. Load Holodeck Program A-3" or some such phrase, and then the Holodeck would load this program and the user would enter. However, we never actually experienced the program itself – only its inputs and outputs. Similarly, in The Matrix, we have a rich embodied experience of human characters that , in reality, are stored inside of a network of fluid-filled pods.

Despite our familiarity and utility with text-based process descriptions, it is remarkable and ironic that a hyper-real environment such as the Matrix affording real-time synthetic interactions and simulacra would have to be programmed by strange-looking rivulets of green rain, which are not obvious to anyone, presumably except for the operator well trained in this postmodern descendant of cuneiform script. This semiotic condition presents a stark contrast: practically unlimited full-sensory simulation on one hand produced by the program, and what amounts to glorified typewriter symbols on the other defining the program itself. It is as if one provides you with a highly maneuverable hypersonic jet plane to fly with the caveat that you need to pilot the plane by tapping on a straight key to produce Morse code dots and dashes. One would expect that, just perhaps, the capabilities that form programs and data might avail themselves of the practically unlimited human-computer interface that the Matrix provides. Rotman (2000, p. 67) poses the question that forms this concern, "What if language is no longer confined to inscriptions on paper and chalkboards but becomes instead the creation of pixel arrangements on a computer screen?"

## 26.3 AESTHETIC COMPUTING: TURNING COMPUTERS INSIDE-OUT

Computers have shrunk in size, and increased in number, considerably over the past half-century. We are familiar with news stories about how ever smaller and thinner computers and software are now ubiquitous in our culture to the point where we carry or wear them in our daily routines. The decrease in size and increase in number creates a situation where computing effects most consumer products. For example, the digital video recorder enables time and place shifting for movies and television shows. What is just as interesting is exploring how computing affects us and our thinking. Turkle (2004) explains this psychological phenomenon and closes with the phrase “we are all computer people now.”

Turkle’s argument has significant ramifications for computing, and I would go one step further to suggest that the way in which our thinking is changing culturally surfaces deep abstract concepts in computing to us as we use these devices: from number, to information structure, to process. Digital watches and video recorders (DVRs) are good examples. Most digital watches are multi-function. These watches contain the ability to act as a way to tell time, set a stop watch, or wake up to an alarm. To use the watch, you have to learn how to navigate a menu by repeatedly pressing a mode button. In each mode, there are sub-functions refining that mode’s interaction. This experience of mode-button pressing directly maps to a fundamental theoretical structure in computing called a *finite state machine* (Hopcroft et al. 2000). It is not just that the finite state machine is embedded within the watch’s silicon, but also that the human wearing the watch becomes aware of this virtual machine’s structure and its components through the experience of using the watch. The state machine *changes how the wearer thinks*, even though the wearer is probably unaware of the formal mathematical notation of a state machine. The watch’s software internals become embedded within our psychology and culture. A similar process occurs within most other household appliances such as the DVR, however, the state machines in DVRs are more complex than in watches – yet to understand how to navigate the hierarchical menus, one has to become fully aware of a new type of thinking (Negroponte 1996). Effects of computing on thought (e.g., neo-millennial/digital native learning styles) have also been covered in the context of learning (Dieterle et al. 2007).

Experience with computing artifacts is a form of information representation, where the definition of "representation" is expanded as a form of interaction, rather than as a static object in the form of a sign. If the raw elements of computing – information, data, and software – are changing the way that we think and entering into our popular culture, it is natural to suggest that aesthetics of these raw elements can and should play a central role in computing. Aesthetics has evolved from the embodied, sensory, definition to a more comprehensive one offered by Kelly (1998), a "critical reflection on art, culture, and nature." Aesthetics within computing results in new interaction modalities for computing artifacts such as formal languages. Given the preponderance of new ways to connect human with computer, there are many opportunities for creative representation. We categorize and study these new ways using the phrase *aesthetic computing*.

## 26.4 WHY AESTHETIC COMPUTING?

Representation targets of aesthetic computing include terms such as data, information, software, and code. I use these terms somewhat interchangeably because of semantic overlaps. Data can be atomic or in the form of a structure. Code usually refers to software which encompasses both data as well as process. Information theory tells us that all of this is a form of information since information can be decoded as atomic, structural, or procedural. I prefer terms such as code, software, or information when referring to the "computing" part of aesthetic computing since these terms encompass broader categories of items that can be represented, whereas the term "data" in common parlance tends to denote non-procedural forms of information. The argument for aesthetic computing involves emerging areas of computing which have changed:

- Our relationship to each other and to nature. These aspects include ubiquitous (Greenfield 2006, Gershenfeld et al. 2004) and pervasive computing, customization and personalization of interfaces, and the new modalities for human-nature interaction as mediated through computing (e.g., the virtual reality continuum spanning physical, virtual, and augmented reality). Shared and customized interfaces for information visualization (Viegas et al. 2007), code sharing (Reas and Fry 2007), assisted with "remix culture" (Lessig 2008) create a networked, customized (Pine 1999) representational space.
- Our thought patterns, allowing computing artifacts such as information and software to permeate our experience. Salomon (1990) makes an argument for computing changing thought, resulting in *cognitive residues* from human-computer interaction. These studies are consistent with Turkle (2005).
- The importance of experience in computing in human-computer interaction (HCI). Cockton (2011) and Hassenzahl (2011) describe the shift in HCI from efficiency, alone, to experience, and Löwgren (2011) emphasizes the importance of interaction – a core aspect of experience. The emphasis on experience is related conceptually to embodiment (Lakoff and Nunez 2001, Johnson 2007) as a basis for cognition. The relevance of aesthetics in HCI is discussed by Tractinsky et al. (2000) and Norman (2004). Dourish (2001) lays out a philosophical foundation for embodiment in HCI through its beginnings in phenomenology.
- Our need as computer scientists to interact more frequently with artists and designers since they represent the creative component of aesthetic inquiry, and so experience-based representations for the diffusing computing artifacts need to be studied

with the help of artist-scientist collaborations (Buxton 1988, Malina 2011).

## 26.5 HISTORY OF THE AESTHETIC COMPUTING FIELD

I have been teaching a course in aesthetic computing since 2000, and information on the most recent course can be found in (Fishwick 2012). A preliminary paper was published on the concept (Fishwick 2002). A Dagstuhl seminar on Aesthetic Computing (Fishwick and Bertelsen 2002) was co-organized in Germany (Dagstuhl 2011) by myself, Roger Malina, and Christa Sommerer during the summer of 2002. This interaction resulted in several publications (Fishwick et al. 2005, Fishwick 2006, Fishwick, 2007a, Fishwick 2008b). Kelly et al. (2009) represents the most recent published workshop in the area. The use of the word "aesthetics" and "programs" can be found in several contexts, including Mohr (2011) and Nake (2009) who were early investigators in the aesthetics of interaction through the use of computer programs as a means of artistic expression. Knuth (1992) developed literate programming and made note of the importance of aesthetics in programming. Knuth's interest in aesthetics went beyond the purely cognitive, and included artistic forms of typography and layout design for programs. For Knuth, it would seem that computing was an embodied experience.

Aesthetic computing is unusual in that aesthetics is intended to be applied to computing rather than in the inverse direction: using computing to create artistic products. Examples of aesthetic computing, therefore, capture a kind of "boomerang effect" where elements of computer graphics, ubiquitous computing, and mixed reality interfaces can be used to interactively represent that which formed these technologies – namely the information and software.

In terms of academic curricula, Aesthetic Computing has been taught for a decade at the University of Florida in the form of two classes, which are usually combined: CAP 4403 (undergraduate) and CAP 6402 (graduate). The combined classes began as part of the Digital Arts & Sciences (DAS) programs (Fishwick 2012) designed and developed to connect computing with the arts. The class has undergone several stages since 2000:

- (2000-2005) Representational alternatives to software artifacts – from numbers and expressions, to data structures and programs. There was one physical project, with the other two projects resulting in digital representations. The physical product was exhibited in several gallery areas on and off campus allowing passers-by to comment and explore.
- (2006-2009) Alternative representations for mass media and communications. This emphasis required students to employ representational creativity, but with the idea of starting with a contemporary news story and then mining this issue for the software artifacts in the story that were to be represented. The physical project was eventually dropped since many of the computer science students in the mid to later years had minimal design and art backgrounds.
- (2010-2011) Representation using web mining and APIs. This was an effort to create more automation in the representational process with students finding sources of information and then, mostly using APIs, to translate this information into creative representations. Most students use data as their information, but others used more complex web structures (e.g. XML) as sources.
- (2012) A focus on representation of data structures, mathematical models, and dynamic models and programs. The end product is either an interactive game or video production whose goal is to facilitate education of computing concepts for early-age groups and

non-computing specialists. This is the current incarnation of the class (Fishwick 2012).

We use the term aesthetics in the spirit of Kelly's definition, but also extend the concept of "critical inquiry" to include the creative aspect of design and art. This is only natural, for engaging in critical inquiry presupposes and requires the creative act. Studies in aesthetics are numerous (Audi 1999, Kivy 2004) often with underlying attempts to find universal attributes of beauty (Scruton 2011). My view on aesthetics is one that focuses on that which is generated as a result of cultural inquiry, which is to say the vast diversity of design and art forms. This "aesthetics as diversity" approach is similar in spirit to Hogarth (Burke 1943) with the associated phrase, "unity in variety."

## 26.6 TOWARD SOFTWARE AS EMBODIED EXPERIENCE

### 26.6.1 Introduction

Partial justification for the use of embodiment as a form of representation is based on educational learning styles (Dede 2005). Also, our ongoing research indicates a significant correlation between presence and memory in a virtual environment (Fishwick et al. 2010) with results currently in the journal submission phase. Recent mixed reality memory studies such as (Ikei and Ota 2008) indicate positive effects on memory in an augmented environment. Instruments and studies on memory performance within virtual environments are being continually refined and investigated. Parsons and Rizzo (2008) introduce a test of validity for a virtual environment cognitive instrument called VRCPAT. Johnson and Adamo-Villani (2010) note significant effects of immersion on short term spatial memory. Embodied interaction with technology provides us with an understanding of internal logic, software, and process usually through pure experience. For example, we learn the state machine of a DVR through repetitive DVR use. While a large population may require this learning, not everyone may be required to take representation to the next step: from interaction to reflection and reification. The latter steps, however, have potential utility in entertainment (the arts, games) as well as in education.

### 26.6.2 Audiovisual Explorations: Steampunk Obesity Machine

Let's consider one such artifact, which is defined by a system dynamics model found in systems science and simulation. Figures 2 and 3 are two different representations of a System Dynamics flow graph (Forrester 1991) capturing the temporal nature of human metabolism.

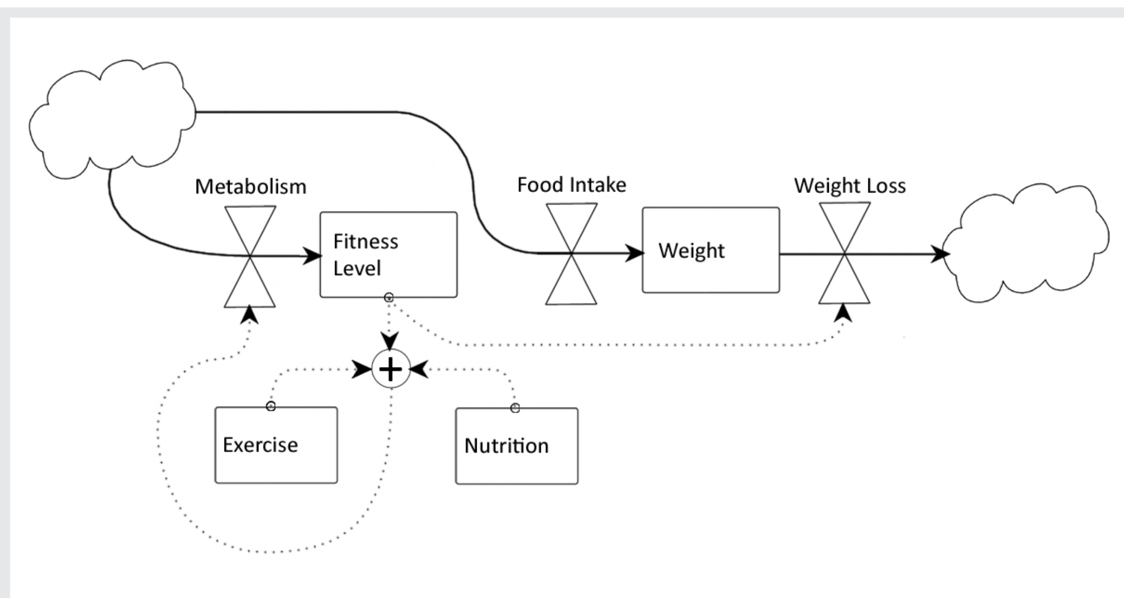



Figure 26.2: A System Dynamics flow graph with two levels (i.e., stocks) and three rates.



The diagram in Figure 26.2 represents a virtual machine based on the analogy of fluid flow. Fluid starts from source node (left-most "cloud" icon) and proceeds to flow through a system of levels separated by rates to a sink node (right-most cloud icon). More generally, the fluid flow can be construed as a kinetic energy flow since fluid velocity is the dominant flow variable. At the start of the machine, at the left, fluids pour into metabolism and food intake to suggest that the more energy, the higher the *Fitness Level*, but also the higher the *Weight*. The rate variable, *Metabolism*, is proportional to a functional combination of Fitness Level, Exercise, and Nutrition. The nature of this precise formula is not present in the model since the model is an abstract representation of the dynamics. The solid curve arrows reflect fluid flow through the system, and the dashed curve arrows reflect control settings to change the rates on the valves. Figure 26.2 is a hypothetical example, and is not put forth as an accurate or valid simulation model of nutrition, but rather to demonstrate that similar diagrammatic models are widely used in science and engineering. These types of models were originally implemented as physical, analog computers although their more frequent existence today is as digital models with a diagrammatic front end authoring capability. The MONIAC, or "Phillips Machine," is one such example (Swade 2000, Ryder 2009) from the analog computing era.

Figure 26.3 shows the same model which is a synthetic rendition of Figure 26.2, reified using a "steampunk machine" since its structure is reminiscent of the cyberpunk aesthetic that continues to be popular since its inception in Gibson's work (Cavallaro 2001). Steampunk culture has connotations of "reclaiming tech for the masses" (Grossman 2009). Water is pumped using steam-power underneath the wooden floor. This water shoots out of two brass orifices that represent the two valve-icons in Figure 26.2. Water filled glass containers represent the level quantities, and wood/brass control rods connect everything together as in Figure 26.2. The human avatar on the left is demonstrating the machine in action to us, or we may become the avatar. The natural question is why anyone might want to construct such a machine when Figure 26.2 might do. For the answer to this, we have additional questions to ask, with possible use-cases.

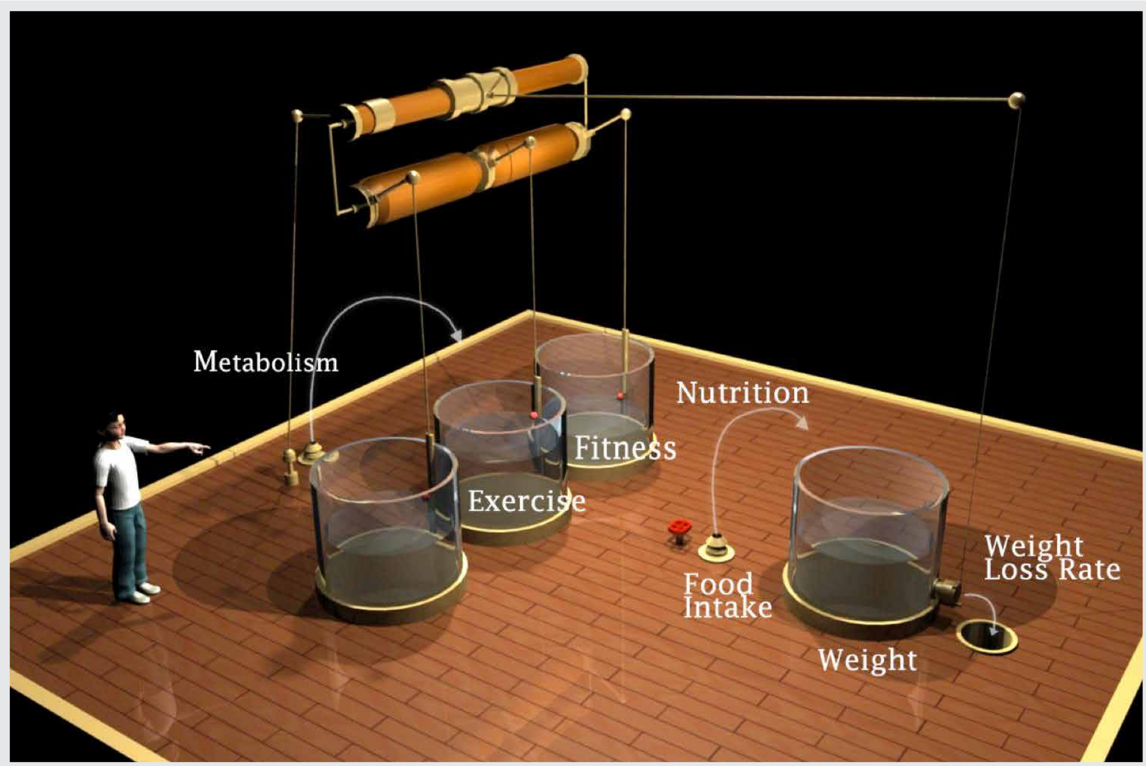



Figure 26.3: A steampunk obesity machine isomorphic to Figure 26.2.



Figure 26.2, and the equations that map to this diagram, are most often used by scientists familiar with the system dynamics method. It is unlikely that these scientists have any interest in structures such as Figure 26.3 mainly because they are comfortable and familiar with more formal representations. However, the vast majority of the population may require additional motivation if they are to understand, and be motivated or influenced by, the more formal representations. Therefore, the machine for Figure 26.3 is appropriate for education and entertainment. It is easy to imagine the machine in Figure 26.3 being engaging especially with game-like features that required certain goals such as stabilizing the water level in the Weight container.

### 26.6.3 Visual Representations of Data

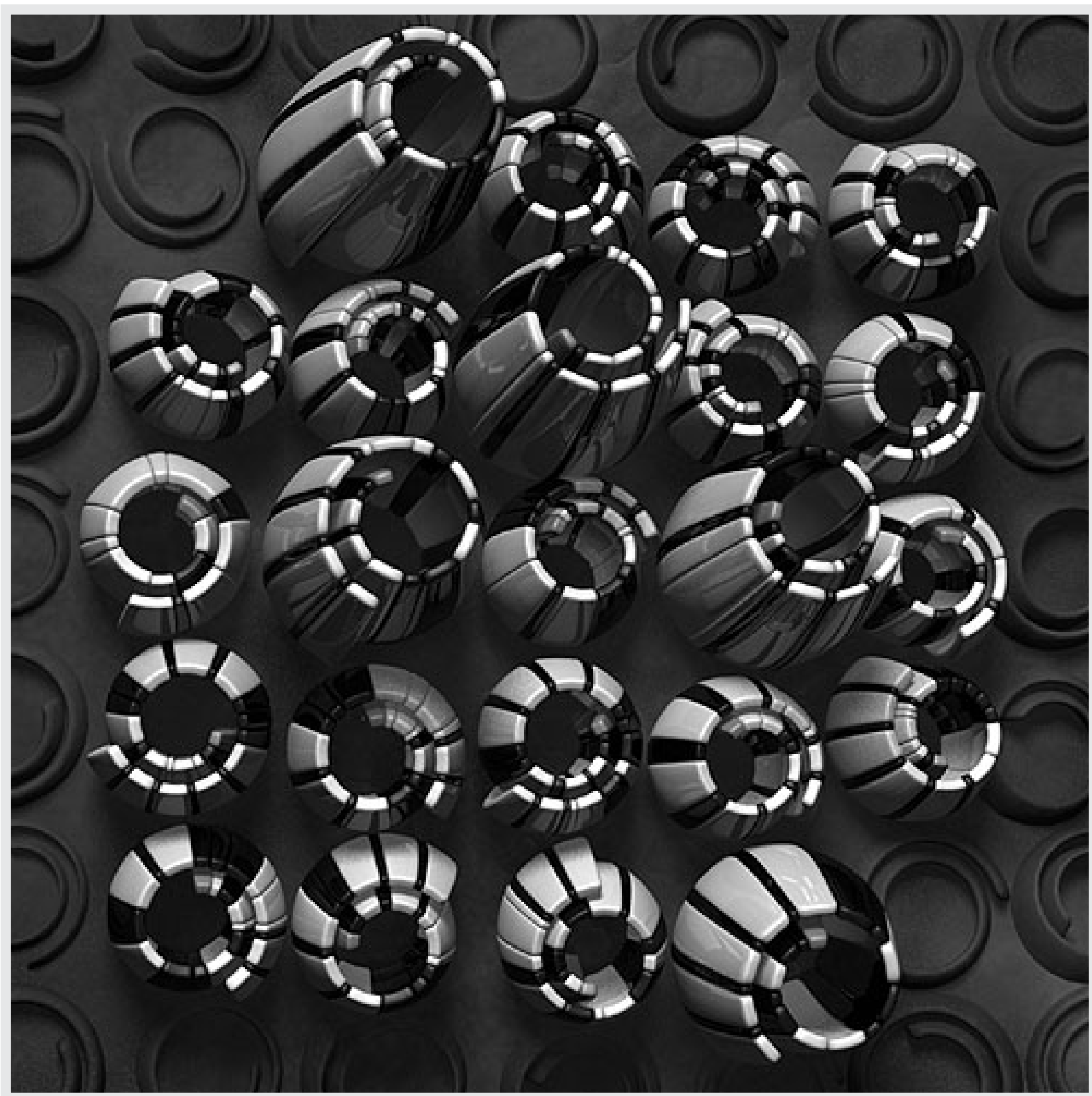
There are numerous additional examples of artworks that, if used as guidance, can lead to aesthetic computing products useful for education. The vast majority of examples are the encoding and presentation of data rather than of program or model. It is logical given that data repositories and accessibility are expanding rapidly and that they represent the simplest and easiest to grasp forms of information. Consider the model of a single number shown in Figure 26.4.

FIGURE 26.4: The relative size of the U.S. debt if it reaches 15 trillion dollars. The large rectangular block represents stacks of one hundred dollar bills.



This encoding of number as a stack of one hundred dollar bills is given context by familiar objects whose size is known through pictures or experience (e.g., the Statue of Liberty, a football field, a truck). One might take this same approach to representing other analog representations of monetary amounts through choosing different familiar objects. A participant's engagement can have both artistic and mathematical consequences. For example, we can imagine performing operations on numbers in this type of representation much as we have done manually in the past with quipus and abaci.

Consider Huff's prime number series (Huff 2006) with two example encodings of prime factors shown in Figure 26.5.



*Copyright © Kenneth A. Huff. All Rights Reserved. Reproduced with permission. See section "Exceptions" in the copyright terms.*

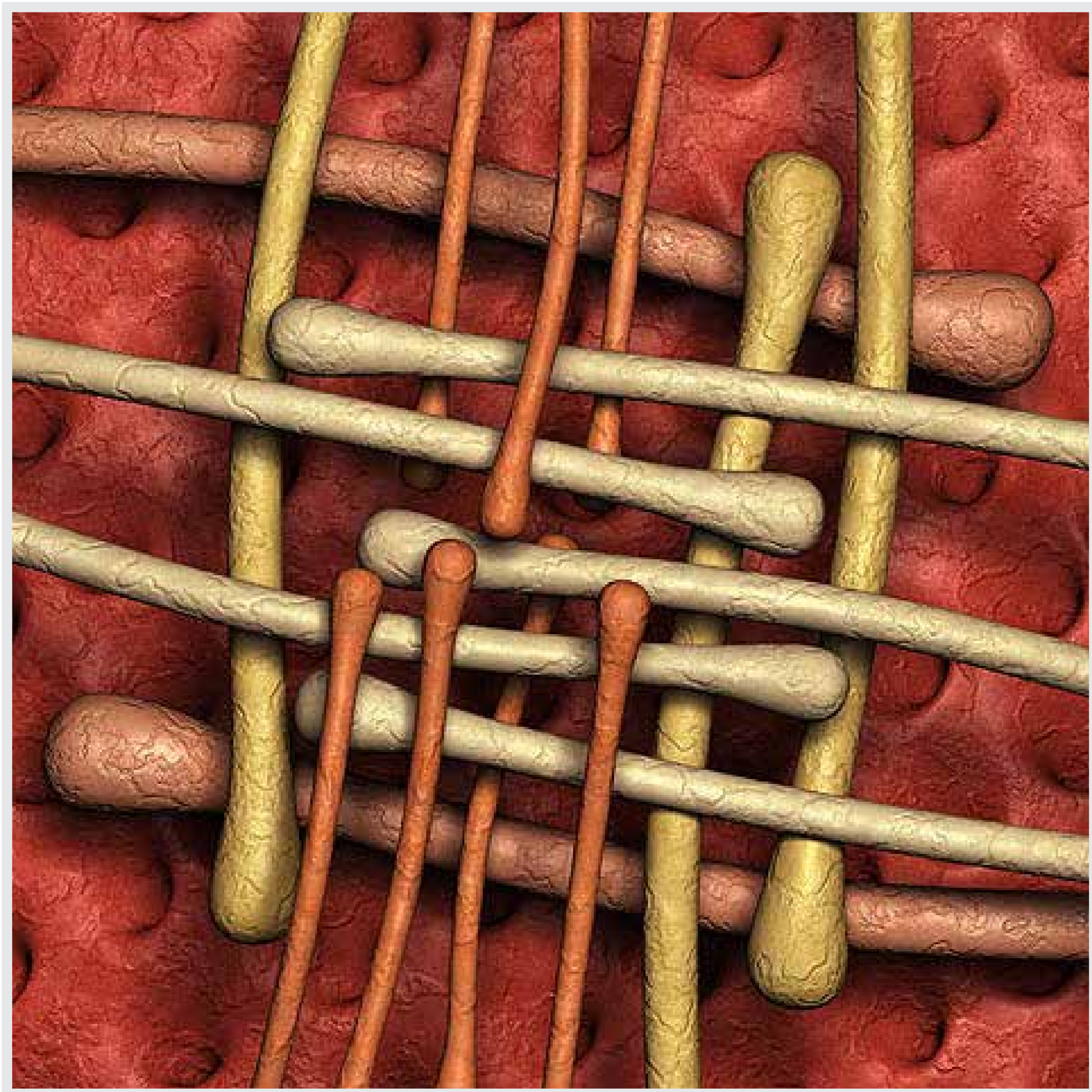




Figure 26.5 A-B: Prime number factorization encodings (EPF: 2003:V:A:997141 and 2000.24).

The two encodings in Figure 26.5 are pieces of fine art, but could also be potentially used to motivate students to appreciate prime factorization through puzzle-making. For example, consider where one might provide to someone a visual encoded integer and then ask that person to identify the number and factors. Figure 26.6 shows two additional examples of information presence: Levin's infoviz graffiti for data, and Living Light. The graffiti is a deliberate mechanism for surfacing numbers of societal relevance in public places. Living Light is a permanent outdoor pavilion in Seoul, South Korea. The pavilion's purpose is to allow spectators to visualize environment levels such as air quality. As pervasive computing extends into the future, most flat surfaces become display surfaces opening up numerous possibilities for bringing information into our daily lives. Figure 26.7 shows a model of a city which is turned into a computer program-like artifact, or automaton, whose output is a musical score.

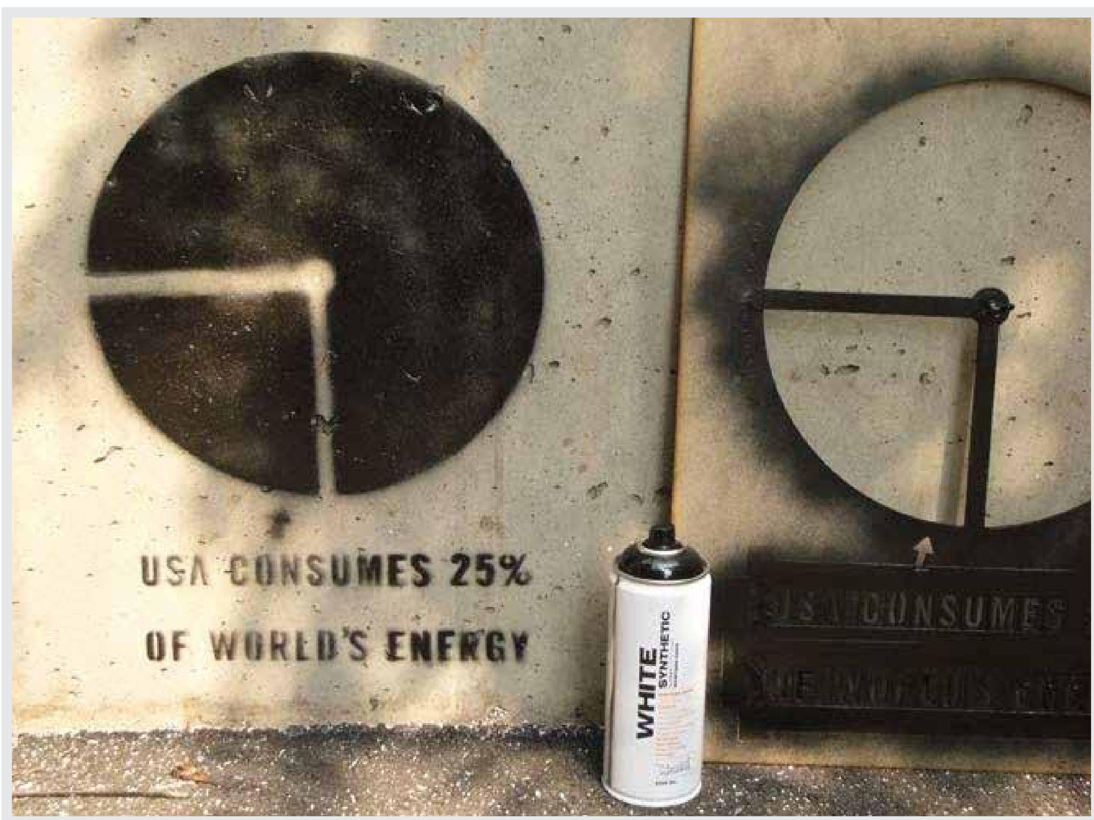




Figure 26.6 A-B: Leftmost: Infoviz Graffiti/Adjustable Pie-Chart Stencil by Golan Levin. Rightmost: Living Light by David Benjamin and Soo-In Yang.

Figure 26.7: Pianola City Music.

### 26.6.4 Textual Representations

The examples so far have been mainly visual; however, it is often desirable to use thinking similar to that described earlier for mathematical expressions, but to extend this to software for example. Emerging areas in the humanities such as software studies (Fuller 2008) and critical code studies (Marino 2006) situate the need for studying formal languages using some dimension of hermeneutics. These areas also provide opportunity for creation of new human-computer interfaces. For example, we may treat software as a full hyper-mediated structure (Roth et al. 1994, Anderson et al. 2000). One can then, through an embodied approach facilitated by link interaction, treat formal language-based constructs as hypermedia.

## 26.7 ART AND DESIGN AS CREATIVE INFLUENCES FOR EMBODIED FORMAL LANGUAGES

The provided examples including Figures 3 through 7 are related to aesthetic computing in different ways. Since aesthetic computing is embodied in formal languages with an educational goal as a final end product, I will overview how these examples might achieve that goal. Table 26.1 contains 5 columns: column 1 refers to a previously described image or product; column 2 is the original medium; column 3 is a hypothesized goal for the last 5 rows (i.e., since the original intention is not known but assumed); column 4 is an example repurposing of the original product for a formal language goal (column 5). Let's consider the 3rd row. The product has been designed to a highly compelling and attractive display of the national debt. This creative use can be recast as a new way to learn *number sense*. The formal language products are only examples and have not been constructed by anyone, however, the original art and designs are dually inspirational – for their original goal or purpose, and for a form that leverages their embodied characteristics for the purpose of formal language instruction.

| Example | Original Product | Aesthetic Goal (hypothesized) | Formal Language Product (example) | Formal Language Goal |
|---|---|---|---|---|
| **Personal Experience (Arithmetic)** | Typographic Image | To illustrate elegance of the mathematical form | Game with moving operators and operands | To teach laws of arithmetic |
| **Steampunk Obesity Machine (Figure 26.2)** | Raster Image art work | To create steampunk genre-related imagery | Video illustrating functional mechanism and control | To teach System Dynamics Methodology |
| **US National Debt (Figure 26.3)** | Raster Image art work | To illustrate the magnitude of the US debt using scale | A tactile set of blocks and objects | To teach number sense |
| **Prime Number Factorization (Figure 26.4)** | Raster Image art work | To celebrate organic forms using prime number encoding | An adventure game using encodings as 3D puzzles | To teach about prime numbers and factorization |

| | | | | |
|---|---|---|---|---|
| **Infoviz Graffiti (Figure 26.5a)** | Graffiti and Template in Outdoor location | To present societal information to the public | An alternate reality game (hunting for graffiti) | To teach concept of percentage |
| **Living Light (Figure 26.5b)** | Outdoor sculpture | To present environmental information to the public | A kinetic sculpture | To teach data structures |
| **Pianola City Music (Figure 26.6)** | Indoor kinetic art work | To explore an architecture-music interface | Indoor kinetic object | To teach concept of a data search via sound |
| **Hyper-mediated software engineering** | Web-based computational literature | To represent cultural knowledge | Hyper-mediated software/code | To encourage learning of how to code |

Table 26.1: An aesthetic transformation to formal language learning objectives.

Table 26.1 portrays aesthetic computing through repurposing existing art works, but this procedure is optional. Formal language-based products that capture the essence of embodied interaction can be designed directly from initial design, to detailed design, and onto an implementation. The Steampunk Obesity Machine (Table 26.1, Row 2) is a case in point. Even though a poster board image (Figure

26.3) was part of a curated art exhibit (Harn 2011), the image was meant as a preliminary design for a virtual machine to teach System Dynamics concepts. The machine has not yet been constructed.

## 26.8 EMBODIED COMPUTING USING SERIOUS GAMING

The discussion of aesthetic computing and the interpretation of it via embodied formal language would be incomplete without reference to video and console game cultures. Two examples are illustrative: logic circuits in the game Minecraft (2011) and the game called Code Hero (PrimerLabs 2011). Minecraft is a "block game" where players move around a space and build blocks using a mining metaphor. Some of the procedural capabilities within the game have engaged members of the community to create basic circuits, leading up to full-fledged computers out of the logic circuitry. Since Minecraft is highly interactive, and invokes a sense of presence to boot, this type of hacking is consistent with the concepts in aesthetic computing: players are working together to form circuits through embodied interaction. Primer Labs recently created a game called Code Hero where the play learns a programming scripting language such as Javascript. It is the means for this pedagogy, however, which places it squarely in the embodied realm: a player has a gun that "shoots code" at a target object, thus causing that object to react to the code. This is, in actuality, a reified form of data flow in a manner similar to the capabilities within lambda calculus and languages based on that formalism such as Lisp (e.g., consider the "map" functions where a function can accept another function as input and then apply that function to arguments, producing output). Figure 26.8 shows a Minecraft arithmetic logic unit (ALU) described by Ganapati (2010), and Figure 26.9 shows a snapshot from Code Hero.

FIGURE 26.8: Arithmetic Logic Unit built with "redstone" in an immersive play space using the Minecraft game engine.


FIGURE 26.9: The Code Hero Game: talking to the Ada Lovelace avatar prior to entering a space to learn how to script code by shooting scripts at objects.



## 26.9 COLLABORATIVE ROLES, USABILITY, AND EXPERIENCE

Aesthetic Computing begins with a formal language construct such as a number, data, model, or software. Then the challenge is to represent this construct through embodiment. We noted that "embodiment" can be as simple as pure reification without representation of existing objects when we demonstrated the ability to grab hold of numbers and move them toward operators. However, reification can also suggest object representation as in Figures 4 through 7. I need to address the "who" and "why" aspects of aesthetic computing.

First, who is going to be creating these representations? In the case of collaboration, I recommend teams of humanist scholars, artists, and computer scientists. Humanist scholars bring to bear different philosophies and theories which can help shape the resulting representation. The artist has the creative perspective and tools to create the representation, and the computer scientist can serve two roles: to help construct tools used by the humanist and artist in the extraction of information and in enabling the interaction that ensues through externalizing embodiment in the human-computer interface.

Second, who is going to use the representations? Students in my aesthetic computing class are often initially confused why one would construct anything but diagrams. This confusion is expected, but we must be careful when defining usability: usable for whom and for what purpose? We need to identify 1) the goal of the representation, and 2) the end target users. Goals for the embodied representations are education, arts, and entertainment (e.g., cinema, visual and performing arts, fiction). Target users may be any grade level in school or some segment of the general public. From a psychological perspective, a broad view of "usability" can encompass user goals including: increased valence, motivation, and attitudinal change, as well as improved short or long term memory. Mathematicians and computer scientists are not the target, as these populations are adept at using existing notations. Aesthetic Computing is less stressed on information extraction and more on the use of entertainment, arts, and humanities on formal languages with the largest practical effects being in education. Thus the target users are formal and informal learners of all elements of formal language-based instruction (e.g., mathematics, computer science).

The roles of participants in aesthetic computing will likely be different given the interests of each party. For the computer scientist, for example, Figure 26.5 serves as a design template for the creation of special effects and interactive games for the purpose of expressing elements of prime numbers and the factorization process into these numbers. The artist's work is a medium through which this aspect of formal language is creatively expressed. The goals of the artist and computer scientist are clearly different, but the means (i.e., representations of prime numbers) are common. This difference in ends, with similar means, plays out in the other examples. For instance, Perl poetry (i.e., poetry created using the programming language, Perl) may be an aesthetic product to the writer – a valid end in itself. To the computer scientist, this product represents a medium in which to express a different end – the formal language "message." Therefore, aesthetic computing by its arrangement of words comprising this phrase is focused on computing – the learning of formal languages. However, aesthetic products play a key role in this learning activity and allow for the artist, scholar, and computer scientist to collaborate with different intentions and goals.

Other areas related to aesthetic computing are information visualization (Card et al. 1999, Ward et al. 2010), and software visualization (Eades and Zhang 1996, Stasko et al. 1998, Zhang 2007, Diehl 2007); however, the goals of these areas are generally quite different than for aesthetic computing. In information visualization, the goal is efficient communication of data and information, whereas for aesthetic computing, the goal is education through highly embodied, and interactive, aesthetic products in the forms of art and entertainment. As such, Aesthetic Computing fosters a deeper experience than building representations meant for immediate consumption (e.g., newspaper diagrams and maps). Readers will observe that the use of metaphor is rich within the high level interactions with computers. We are an interface culture (Johnson 1997). However, the metaphors used on the "desktop," for instance, have not yet made their way into the core of mathematics and computing. Efforts such as computational thinking

(Wing 2006) are a move in the right direction.

Laurel (1991) presciently captures a prerequisite for aesthetic computing in her "Computing as Theatre." However, Laurel was mainly constructing a case for human-computer interaction as a complex theatrical production, involving many of the same elements found in theatre. The *use* of computing, and its associated interaction phenomena, are like theatre. However, what we find is that as we break open the lid of the black box containing the atomic elements of normally hidden data, formulas, code, and models is that computing is theatre all the way down.

## 26.10 TOWARD A METHOD OF AESTHETIC COMPUTING

While it is interesting to pose ideas and directions, a procedural method is something that can help to forge a discipline even if only as a general guide. Fishwick (2007a) was an initial attempt at this process with a small example of code that was represented as a collection of rooms in a building, complete with a partial narrative for context. Figure 26.10 serves as a basis for describing the approach used in (Fishwick 2012):

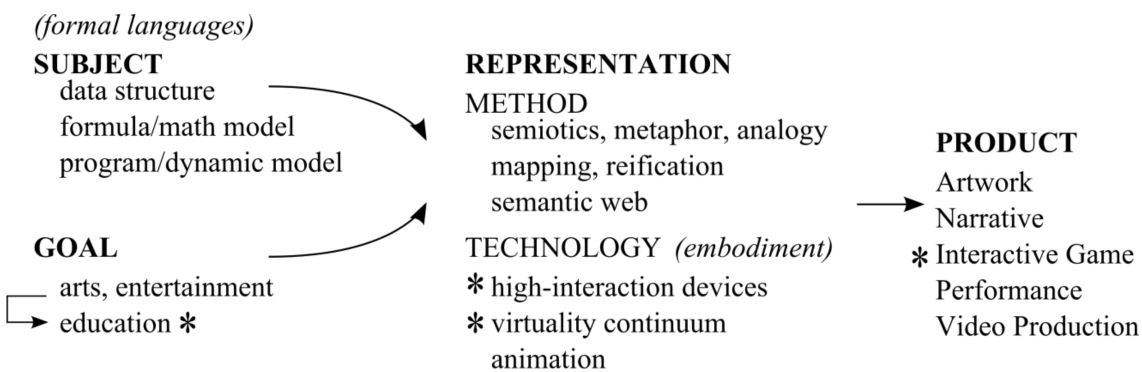



Figure 26.10: Aesthetic Computing Method.

*Courtesy of Paul Fishwick. Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.*

We begin (in the top left of Figure 26.10) with a formal language construct that is to be conveyed to non-specialists in mathematics and computing with the goal of broadening the exposure of computing concepts. The asterisks denote current emphases in (Fishwick 2012). Target users will depend on the type of formal language. If the goal is number sense, and the numbers are fairly simple, we may be looking at elementary school children. If the formal language is simple algebraic formulas, we may be looking at 8th grade mathematics. More complex mathematical and computing structures may require higher grades, including universities and in postgraduate, informal learning contexts. One of the desirable outcomes of this approach to representation, though, is to expose very young children to seemingly complex data structures and programs by using games and video as motivational media. I expect that the approaches may serve as 1) scaffolding for later, more traditional, instruction and notations, and 2) secondary devices (e.g., puzzles) to reemphasize concepts that some learners find difficult using standard notations. The goal is not to eliminate standard notations as this would be counterproductive. Representation is divided, in Figure 26.10, into two components: methods that achieve representation and technologies that support embodiment. End products that emphasize, or surface, embodiment can vary. A good piece of fiction can create a strong sense of presence and virtual embodiment, whereas a weak interactive game may be left ignored if not well designed.

## 26.11 NEW CONNECTIONS

A primary goal of mine in fostering aesthetic computing is to link disciplines – especially those in computing to the humanities and arts. As evidenced by designers and humanist scholars, artifacts such as "code" and "data" are now being interpreted and recreated. There are many reasons for this. Perhaps, the ubiquitous computing trend is the most significant driver – software is everywhere and so, by natural extension, cultural. I welcome the artists, designers, and humanists into the "formal languages" space and hope that through collaborations and interdisciplinary discussions and critique that we might re-humanize core elements of computing, and perhaps even mathematics.

## 26.12 DISCIPLINARY AND TECHNICAL CHALLENGES

The area of aesthetic computing is not without its challenges. The goal is to leverage embodiment theories toward building new computer-based interfaces for learning formal languages. Disciplines that I have covered have sub-areas that are all targeted toward this goal, but significant challenges remain for each area:

- *Mathematics*: the literature in mathematics education, and in the application of cognitive linguistics within mathematics learning, is well-founded and supports aesthetic computing. This body of knowledge, however, is more focused on analysis and theory construction rather than, through analogy, building new interfaces in mathematics education to take full advantage of the embodiment theories through realization. Some efforts in *virtual manipulatives* are a good start, but this work should expand to employ the next generation of interface capabilities that stress embodiment (e.g., multi-touch displays, body tracking, mixed/virtual reality technology).
- *Computing*: the literature in computing education provides fairly easy-to-use interfaces for seeing the results of executing programs; however, the programs are often limited to the canonical alphanumeric notation with all of the human interaction being in the program execution rather than inside the program. Efforts at software visualization move in the direction required by aesthetic computing, and yet, there is a much wider set of possibilities for representation if the goal is to teach non-specialists especially through immersion, situated learning, and interactive games. Diagrams are fine for communication, but if the goal is to explore deeply embodied approaches for learning, additional media and newer interfaces – as recommended for mathematics education – should be more thoroughly investigated.

- *Humanities:* the work in cognitive linguistics, and resulting embodiment theories, ground the work in aesthetic computing, but as with the work in the philosophy of mathematics learning, there is little corresponding effort in realizing these theories in a human-computer interface. Conversely, the work in cultural theory production is recently targeting "code" specifically as a new type of literacy (e.g., critical code studies). And yet, this production tends to avoid linguistic analysis and instead focuses on socio-historical analysis. New, embodied, interfaces for code can build off of the scholarly analysis, but these interfaces should also be informed by key facts of semiotics (e.g., analogy and metaphor) which lie at the foundation of formal languages. There appears to be a bias toward textual notation rather than exploring broader forms of "embodied literacy," which would include textual notation as one dimension.
- *Art & Design:* works of art have traditionally treated formal languages as "black boxes," tools needed to create art or designs. Unlike in the humanities, where code has become subject material, in art, code tends to be treated purely as a tool, whether embedded in package or programmed via a text-based development environment. The only exception to this observation would be in typography within graphic design, where the subject material is the text. More explorations are required so that formal languages become the active subjects of artwork.

Each one of these four areas has some common challenges. Observing that analogy is the engine of metaphor in scientific practice, aesthetic computing products can be created with an increased attention to analogy. Another observation is that with the exception of Art & Design, there is a classical focus on alphabetic notation. Such notation serves us well and has enriched our formal languages. However, there are other types of notations that exercise more of the body's sensorimotor functions. Diagrams are a good place to start in seeing this transition since with diagrams spatial metaphors for text-based notations abound, but we should not limit our embodied explorations to diagrams.

A primary aesthetic computing challenge is technological. It is still relatively expensive to build new interfaces based on the types of products described by the figures previously shown. “3D modeling” as a real-time technical interface capability is nowhere near the futuristic landscapes of Tron, the Matrix, and the Holodeck. Modeling and animating in three dimensions remains a major challenge compared with diagrammatic approaches, and even diagram-based software modeling (e.g., model-driven architecture) struggles for acceptance in the marketplace of software engineering solutions because of the relative ease of using textual symbols. Human-computer interaction solutions are expanding in scope and capability, but we still are a long way from being able to easily and inexpensively become embodied in our formal language constructs.

## 26.13 SUMMARY: THE ARGUMENT FOR EMBODIED FORMAL LANGUAGE

This chapter began with personal experiences in mathematics and then moved on to discussions of embodied cognition, along with some examples of where aesthetic computing could be applied. The area of aesthetic computing rests primarily on the foundation of embodiment – whether we believe that our bodily interactions form our thought. This assumption of embodiment runs deep in philosophy. We all recognize that we have body and mind, and most would agree that the latter is the effect of the former. It is only fairly recently, though, that literature has arisen to indicate a strong relationship to the extent that thought itself, even for abstract objects, is embodied. The theory that undergirds embodiment is compelling, but we have the nagging question about how this theory can change what we do and how we act. If I imagine that I am imagining grabbing and pushing a number through a pseudo-biological membrane during arithmetical operations, I want to build a human-computer interface that reinforces this mental sequence by infusing theory into practice. This perceived need matches the aesthetic computing hypothesis stated at the start of the chapter: *Given the embodied nature of cognition, we should realize this embodiment through novel human-computer interfaces for learning formal languages.*

Achieving this realization involves a more thorough understanding of the interplay among disciplines and how embodiment theories in those disciplines interact and connect. The realization also requires a host of newer “virtuality continuum” technologies that allow us to achieve what Biocca refers to as degrees of progressive embodiment (Biocca 1997). The technologies and their characteristics are overviewed for virtual reality by Sherman and Craig (2002), and by Bowman et al. (2004), and for augmented reality by Bimber and Raskar (2005).

## 26.14 WHERE TO LEARN MORE ABOUT AESTHETIC COMPUTING

For a thorough understanding of computing as a discipline, and its artifacts which are represented in aesthetic computing, the 1998 ACM Computing Classification System (CCS 1998) serves as a good starting point. Even though my treatment of aesthetics is based on its original, perceptual definition, Kelly (1998) collects that which erupted from this kernel in philosophy and the arts in four volumes. Even though information visualization is centered on efficient communication (e.g., reading the equivalent of a diagram in a newspaper), some archives such as *infosthetics* curated by Vande Moere (2011) are broader and contain a wide variety of potential use cases – from efficient communication to experience, education, and play. For text-based representations, HASTAC (2011) serves as a high level repository of bloggers and projects, many of which are associated with *digital humanities*. The reader is encouraged to review articles cited in this chapter.

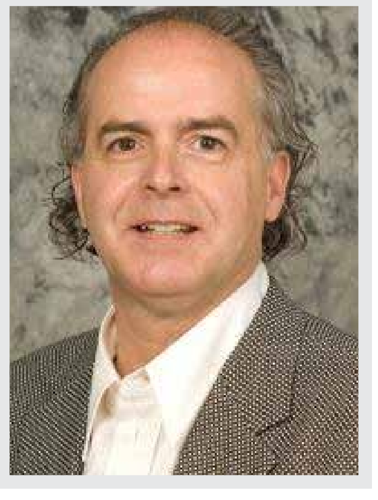
## 26.15 ACKNOWLEDGMENTS

I would like to first acknowledge all individuals who have participated in this journey from its inception, including my colleagues in the arts, natural and social sciences, computer science, and mathematics. Students in my Aesthetic Computing class have had to put up with these ideas, and they have produced wonderful products that I could never have imagined. Thanks to the following colleagues who took time to make very good critical remarks on earlier forms of this manuscript: Sophia Acord (University of Florida), Michael Kelly (University of North Carolina at Charlotte), Mads Søgaard (Interaction Design), and Kang Zhang (University of Texas at Dallas). I take responsibility for any errors and omissions.

## 26.16 COMMENTARY BY MICHAEL KELLY

How to cite this commentary in your report

**Michael Kelly**



*© Michael Kelly*

I am a Professor of Philosophy at the University of North Carolina at Charlotte, and Editor-in-Chief of the Encyclopedia of Aesthetics (Oxford UP, 1998). A new expanded, revised, second edition of the Encyclopedia is forthcoming, and will include entries on aesthetic computing. I organized and participated in a panel discussion on information aesthetics at SIGGRAPH in 2009. And I or...

Michael Kelly
Michael Kelly is a member of The Interaction Design Foundation

> "*The masters of information have forgotten about poetry, where words may have a meaning quite different from what the lexicon says, where the metaphoric spark is always one jump ahead of the decoding function.*"
>
> -- *J. M. Coetzee, Diary of a Bad Year* [1]

## 26.16.1 Fishwick's Encyclopedia Entry on Aesthetic Computing

Paul Fishwick has a well-developed, impressive research and pedagogical platform at the University of Florida from which he's been exploring one of the particular versions of aesthetic computing, although, being involved since its inception, he also has a general sense of the field. In fact, there would hardly be such a field without him. The question I want to ask is how much his particular projects are influencing his general conception of aesthetic computing and whether he's achieved the appropriate, if difficult editorial balance here [2].

Fishwick begins with a pedagogical focus: personal experiences in mathematics that led to discoveries and explorations of embodied cognition. In particular, he analyzes "the aesthetic transformation to formal language," using the concept of embodied knowledge, understood as a perception-action feedback loop based on the idea that embodiment is a form of representation, not just an insignificant step in the process of a strictly cognitive mode of representation. From there he argues, with a rich set of projects, that "The purpose of aesthetic computing is to deliver knowledge and practice of formal languages using aesthetic products as a vehicle." In short, the examples of teaching abstract mathematical concepts that led Fishwick to aesthetic computing have continued to have structural as well as thematic roles throughout his entry and have largely determined his conception of the field. The result is an excellent but partial picture of aesthetic computing that, if taken for the whole, would be misleading.

Fishwick acknowledges a broader conception of aesthetic computing in the "Why Aesthetic Computing?" section (26.4). But he does not adequately clarify it or show concretely how it informs or otherwise relates to his research. Has he perhaps changed his view of aesthetic computing over the last decade? Back in 2006, he made a number of statements about aesthetics that point to a conception of aesthetic computing broader than what is generally evident in his Encyclopedia entry (Fishwick 2006). For example, he says that aesthetics reaches "beyond classic concepts such as symmetry and invariance" and encompasses "the wide range of aesthetic definitions and categories normally associated with making art." Yet now he seems to limit aesthetic computing to often classic concepts specifically relevant to his projects. Quoting the Preface to the *Encyclopedia of Aesthetics*, he embraces the idea of aesthetics as "the philosophical analysis of the beliefs, concepts, and theories implicit in the creation, experience, interpretation, or critique of art." However, when he references this same idea in the Encyclopedia entry, he seems to pull back from its full implications for aesthetic computing. Fishwick also says earlier that aesthetics has logical as well as material aspects, so it can extend to computing as well as art. Taking discrete mathematics as an example, he claims that aesthetic computing encompasses notions of formal language, geometry, and topology, and from such claims he concludes that *aesthetic computing corresponds naturally with mathematical formalism*. In his current research and Encyclopedia entry, Fishwick focuses mostly on this last sense of aesthetics and develops, albeit very well, only the narrower view of aesthetic computing it implies.

Yet, the case for a broader conception of aesthetic computing can be made from within Fishwick's own projects because he argues that aesthetic computing rests primarily on the foundation of embodiment, which is itself a very important research topic in aesthetics and a number of disciplines (e.g., cognitive psychology, affective computing, philosophy of mind, etc.). But even here Fishwick's sense of embodiment is mostly cognitive and pedagogical because it's linked principally to formal languages. This may seem like an appropriate link because computing is so much about formal languages. But isn't the whole point of aesthetic computing to develop and sustain a richer conception of computing? With a richer conception in mind, in effect, the art historian and theorist Caroline A. Jones offers a more art-centric and aesthetics-informed account of embodiment that is focused on the impact of computerized technology on the human body, on the "techno-human." [3]. She begins by arguing that the best way for the critique of our techno-culture to keep pace with "the speed of technological innovation" is "to take up these technologies in the service of aesthetics," which provides "a site for questioning" how our "bodies are interacting with technologies at the present moment." Aesthetics provides contemplative space for such a critique because it "buys us time and space" to encounter and reflect "on embodied experience in an ever more technologized world." That is, aesthetics sets up critique within computing to examine how human-computer interactions impact our bodies. The goal of such critique is not merely to understand all the computer-generated bodily interactions that have been experienced already but to explore which ones could be experienced, and, moreover, which ones we would prefer to experience going forward. In the end, a major advantage of Jones's account of embodiment is that she makes it clear that this kind of critical thinking internal to computing already has a name with a long tradition: aesthetics. By making the links among embodiment, computing, and aesthetics explicit, she offers broader conceptions of computing and aesthetic computing alike.

Jones's account of embodiment, which is explored by a number of artists and theorists involved in the *Sensorium* exhibit or catalog, also dovetails well with the aesthetics of participatory art practices that have developed recently in contemporary art, which would also help to broaden aesthetic computing [4]. Participatory art is, in brief, the convergence of various art forms that emerged in avant-garde modernism or contemporary art: interactive art, installation art, performance art, conceptual art, new media art, public art, socially engaged art, etc. Such convergence has altered the aesthetics of contemporary art in ways (e.g., agency is collective, form is participatory, interactions are transformative) that resonate in computing, too, as it becomes ever more ubiquitous, participatory, collaborative, social, and interactive. Since a central concern about aesthetic computing is how aesthetics is relevant to computing, it would help this cause to examine the most recent developments in the aesthetics of contemporary art. This does not mean that the aesthetics of classical or modern art are not relevant, but since participatory art is emerging in part because of the impact that computing has already had on the production and reception of contemporary art, participatory art is an excellent area to explore while developing aesthetic computing.

In addition to the editorial imbalance, my other principal concern with Fishwick's Encyclopedia entry is that he regards aesthetics primarily as a means ("a vehicle"): "aesthetic computing is embodied formal language with an educational goal as a final end product." As a result, the critical thinking core of aesthetics seems to be lost. For example, although Fishwick identifies some of his own aesthetic norms in 26.7 (only some of which strike me as aesthetic) and his assumptions about aesthetics (e.g., that a principal concern is still the "universal attributes of beauty"), he doesn't analyze them critically [5]. For example, the "unity in variety" concept he endorses is a strategy in 18th century British aesthetics (developed by Francis Hutcheson and others) to identify a property of an object that accounts for its beauty without violating the shared principle among empiricists and rationalists that beauty itself is not a property of any object. How does such a concept or strategy help to clarify the "diversity" of computing or to negotiate between the conceptual nature of aesthetics and the empirical practices of computing? Moreover, as Jones argues, aesthetics can also help to determine the ends of computing by clarifying and critiquing its aesthetic and related norms, so it shouldn't be viewed primarily as the means to achieve ends determined before aesthetic computing was introduced.

To be fair, even if Fishwick's approach to aesthetic computing is narrow in the ways I have described, it may be that the field first has to develop through particular (and thus narrow) projects. Perhaps only then can we initiate a reflective equilibrium between the general field of aesthetic computing and the multivarious, particular projects that Fishwick and others are engaged in. Even though I think the general and particular have to be developed simultaneously from the start, Fishwick has clearly made important contributions to aesthetic computing in this Encyclopedia and his research.

With the same reflective equilibrium in mind, I'd now like to clarify my understanding of a broader conception of aesthetic computing because I appealed to it while critiquing Fishwick [6].

## 26.16.2 Aesthetics in Computing

John Maeda (a computer scientist, designer, and President of the Rhode Island School of Design) once created "Palm Paintings": small, shallow boxes painted in various abstract styles with a Palm computer built into each one serving as its visible center. His stated purpose was to enable us to "think," from the inside, "about what the painting signified." I take it that his point was not necessarily that signification is located materially inside the work of art but, more provocatively, that our critical thinking about the work should take place as the work is being made, if the critical thinking is to be truly inside the work and not merely added as an extra after the fact. The mode of critical thinking here is aesthetics since the key normative issues in art are aesthetic, making aesthetics an integral part of (Maeda's) painting.

In a reciprocal gesture, now imagine that we were to embed aesthetics into the design and production of all the artifacts associated with computers – databases, programs, networks, data visualizations, games, etc [7]. The purpose would again be to think about what they signify and, prospectively, what else we might want them to signify in the future (as well as what other effects besides signification we would like to see). The computing artifacts with embedded aesthetics could be marked in some way to distinguish them from others. We could then hope to learn about ubiquitous computing from the inside, as it is being developed, not merely when it is already being used by people in society.

This reciprocal gesture is not imaginary because, as Fishwick has established, there's been an "aesthetic turn" in a number of areas of computing, leading to the introduction of new subfields such as aesthetic computing, computational aesthetics, database aesthetics, digital aesthetics, information aesthetics, network aesthetics, or software studies [8]. The diverse names, introduced by collaborative research teams of computer scientists and others (e.g., artists, philosophers, art historians), are distinguished by where or, in the spirit of Nelson Goodman, *when* aesthetics is introduced into computing [9]. That is, if we think of the computer stack, the various layers of computing (with bits and hardware at the bottom and user interaction at the top), the choice of name here is a function of when aesthetic norms first enter computing. If aesthetic norms are involved in structuring databases, for example, then we have database aesthetics; if they influence how we give form to information, then we have information aesthetics; if they're part of how we organize networks of people participating in various social media, then we have network aesthetics – and so on within the layers of the computing stack. The lower the layer on which aesthetic norms are implicitly present, the greater the ripple effect the critique of these norms will have on the higher layers of computing [10].

In this light, “aesthetic computing” is the one name among all the options that, in principle, encompasses the entire computing stack and thus best captures the full breadth and depth of the “aesthetic turn” in computing. In exploring more what aesthetics adds to computing, I want to emphasize that aesthetic computing is not merely about the aesthetics *of* computing (merely the design of programs or products, or merely an external critique of the aesthetic norms of computing). Following Maede’s “Palm Paintings,” what I envision is aesthetics *in* computing, albeit with an anticipatory eye to its ethical and social-political impact rather than only its internal structure (i.e., not merely computational aesthetics).

### 26.16.3 What is Aesthetic Computing?

Aesthetics is critical thinking about the norms, concepts, values, or principles guiding or emerging from the production, experience, or reception of art, culture, or nature. Besides referring to the range of theory and practice associated with computer programming, databases, computation, software, operating systems, and hardware (everything from digits to gadgets), the term “computing” (as distinct from “computer science”) captures the recognition that computer science operates in a broad social (moral-political) context. Aesthetic computing is a preferred way to operationalize this recognition because it is critical thinking about the complex set of norms shaping all layers of computing that are, in turn, shaping this moral-political-social context [11].

Despite all the various names for aesthetic computing, there is a common thread running through all the versions or iterations of it. The thread is the recognition among people involved in computing that there are aesthetic norms implicit in the decisions or judgments made on all layers of computing. Accordingly, the main tasks of aesthetic computing are (1) to identify the genealogy and current status of the largely implicit aesthetic norms of computing and to render them explicit; (2) to critique the aesthetic norms with an eye to their moral-political-social implications for users; and (3) to help make decisions or judgments in the future about which aesthetic norms to abandon, revise, or sustain in computing, given (1) and (2), and of course given the technical norms within computing.

We can get a clearer picture of the need for aesthetic computing and its tasks by considering Zadie Smith's review of David Fincher's film, *The Social Network*, and of Jaron Lanier's book, *You Are Not A Gadget: A Manifesto* [12]. What the film and book have in common, on Smith's analysis, is the claim that "Different software embeds different philosophies, and these philosophies, as they become ubiquitous, become invisible." Not only is software not neutral, and not only are there important norms embedded in it (e.g., personhood, privacy, sociality), but software in use also enacts these norms (i.e., puts them in practice). Software does not merely copy our existing norms about ourselves and the world, however, it also enacts new norms and, in doing so, makes a world. The problem, as Smith sees it, is that these invisibly embedded and enacted norms are not discussed critically in advance; rather, they are embedded and enacted by the programmers, most strikingly in the case of Facebook because 800+ million users have had little or no say about its norms. Smith rightly points out that there are ethical issues involved in this case: Why these norms rather than others? Why this format for connecting people with one another rather than another format? What is the quality of the connection? Why this privacy policy? For example, users are expected to give up their privacy to a large extent, and they seem to do so willingly, albeit while reducing themselves to fit the software they are using, according to Lanier, so much so that their "life is turned into a database."

Smith's analysis is relevant to aesthetic computing not only because she points to the *invisibility* of the norms governing Facebook, the Internet, or the Web, but also because when she develops her critique of these norms, she often refers to their "look" or "feel." For example, while we "know" that it is a mistake to believe that computers can personify human relationships, we know this instinctively only by "feeling" the affective consequences of this mistaken belief, which Facebook embodies: "We know that having two thousand Facebook friends is not what it [friendship] looks like." What is this look that we feel and that enables us in turn to know that certain norms embedded and enacted in Facebook may be problematic? We come to learn *that* Facebook is doing something to us through the invisibility of its underlying norms and, if our continued critical reflection is successful, we'll come to learn *what* Facebook is doing to us and, moreover, whether there are any alternatives. To succeed, we need to render visible the invisible norms operating in Facebook so that we'll have "a good reason" for at times feeling "discomfort at the world they're making [in Facebook]." This kind of critical thinking is precisely what aesthetic computing offers because one of its main tasks is to render explicit the implicit norms of computing.

But let me return to the question: Why aesthetics? We might first ask, why philosophy? Smith answers this second question by emphasizing that "it's the *idea* of Facebook that disappoints," not merely the implementation of its idea. To analyze its idea, we need philosophy to counter what she sees as a general cultural tendency in the Anglo-American world to "race ahead with technology and hope the ideas will look after themselves." We need to examine the idea of Facebook and all the other ideas enacted on the Web and Internet before, in Lanier's words, we become "locked in" them, or "entrapped in somebody else's careless thought," which means that we are locked into the invisible norms shaping these ideas and, once those norms are enacted on the Web or Internet, shaping our world and us. But why turn to aesthetics in particular to examine these ideas involving ethics (e.g., security), metaphysics (e.g., personhood or virtual reality), etc.? Returning to Smith's discussion of the "look" of Facebook, and remembering Jones's account of embodiment, the closest we come to experiencing the invisible norms that are enacted in software on the Web or Internet is by experiencing the affects they create on us, the users. Many of these affects are visible, but they involve all the senses (hearing and, increasingly, the tactile), just as works of art do and just as our aesthetic experiences of everyday life do. Aesthetics brings the affective dimensions of our experiences of computing to the fore, and it does so in a way that provides a basis for critique of the sort that Smith, Jones, and Lanier are exploring. These critiques are examples of aesthetic computing in action.

To take another kind of example clearly internal to computing, there has been an "aesthetic turn" in the area of human-computer interaction (HCI) because some researchers believe it is important to obtain a fuller picture of the "user" now that computer interfaces are more interactive, participatory, immersive, and ubiquitous [13]. In a word, they need to understand the user in affective, moral, and political as well as cognitive terms in order, in turn, to create the right (i.e., effective, usable) interfaces. So aesthetics comes into the picture as the notion of usability becomes normatively more complex. Why turn to aesthetics? A major reason is that aesthetics has a long history of critiquing the particular kinds of affective and cognitive interactions and modes of participation constitutive of our experiences of art, and these critiques are relevant to the critiques of the affective-cognitive experiences of the user in human-computer interactions [14]. These interactions (with their own modes of participation) also have moral and political dimensions because users have to be treated fairly (e.g., in matters of access, whether for economic or disability reasons) and their political or cultural beliefs have to be respected. Here, too, aesthetics has a history of critiquing works of art in relation to moral-political as well as aesthetic considerations. The aesthetic turn here, whether in HCI or in any other field of computing, is therefore not a narrowing of moral-political-social impact to aesthetic questions; rather, aesthetics provides a philosophical structure for thinking critically about norms that are moral, political, social, and aesthetic at the same time [15].

## 26.16.4 Aesthetic Computing and Science

Although the critique of aesthetic norms in computing with an eye to their moral-political-social impact is a relatively new process, we can more easily appreciate its relevance and importance if we see it as an augmentation of the existing practices of critical thinking in computing [16]. That is, computing has always critically analyzed its normativity, even if the norms have been understood mostly in technical terms (e.g., what is most efficient or effective). The emergence of aesthetic computing stems from the recognition within computing that its norms are more than technical, as we saw in the case of HCI. So aesthetic computing is principally an outgrowth and refinement of the recognition of the complex normativity always already operative within computing. This is an important point to emphasize because some computer scientists may view aesthetics the way they at times view ethical, political, or other issues seemingly external to computing: those issues are not relevant to what they do qua scientists (given their methodologies, aims, etc.) and thus to give such issues methodological credibility can only place constraints on science. However, if aesthetics (and the related normative) questions are understood as emerging from within computing, scientists no longer need to be concerned that aesthetics is constraining computing.

Yet researchers may still worry that aesthetic computing will change computer science in ways that would make it less scientific, especially if Roger Molina is right that the strong claim of aesthetic computing is that it will generate new objectives that “would not naturally have evolved within the computing sciences” and, moreover, that will “redirect the future development of computing.” [17]. That is, the transition from implicit to explicit aesthetic norms on the layers of the computing stack may have the result that we will change technical as well as aesthetic norms and then change the objectives of computing on that basis. But, again, if computing is normative and the self-critique of normativity is part of science, the only real change resulting from aesthetic computing is that the aesthetic norms always already part of computing will now be explicit and critically examined. How can computing not benefit from more self-critique, since the revision of its internal norms is part of the engine that has driven progress in modern science, on its own terms? For example, as computing becomes more conscious of the design issues that could contribute to environmental sustainability, that may change certain objectives of computing but it would not make computing less scientific, for if it were to become less scientific, it could not contribute to sustainability. In short, aesthetic computing shows how seemingly external norms are actually internal to computing.

The issue of the status of computing as a science is worth dwelling on even longer because it can stop the discussion of aesthetic computing cold. Some may still worry that aesthetics involves taste and is thus subjective. In this light, to integrate aesthetics into computing would be to introduce subjectivity into an otherwise objective science. However, what is actually happening here is that computer scientists are recognizing (a) that the normative complexity of computing has already shaped their idea of science, making room for a more interdisciplinary approach to computing, and (b) that computing is more than a science, not only because its moral-political-social impact entails too many nontechnical issues that scientists need to understand in order to develop computing internally, but also because the implicit nontechnical norms of computing are already shaping its development in ways that need to be analyzed critically for the sake of computing – as well as for our sakes as we live and work with computers [18]. In short, the aesthetic turn in computing is a way to critique its nontechnical norms in order to strengthen its status as science at this stage of its development. Why aesthetics? Again, because it is a long-standing field of philosophy that has developed a variety of ways to think critically about aesthetic norms as they are related to moral-political-social as well as technical norms.

### 26.16.5 Open Aesthetic Properties and Objects

Now, if aesthetic norms are always already a part of computing, why is aesthetic computing barely a decade old, though aesthetics has its origins in the eighteenth century and computers have been around for decades already? One explanation, according to Fishwick, is that computing had to develop to a certain stage before its connection to aesthetics could clearly emerge: "We have had to wait for the technology to become available to leverage the arts," especially in the fields of HCI, ubiquitous computing, augmented reality, and virtual reality [19]. Yet if aesthetics is so obviously relevant to computing, why was this delay necessary? After all, aesthetics is a form of critical thinking and computing has relied on critical thinking to evolve, so why didn't aesthetic computing emerge earlier? Another explanation why its emergence has been slow, besides the worries about science being constrained or becoming subjective, is that too many people in computer science seem to have rather narrow, sometimes outdated ideas about aesthetics and thus have not been able to see its relevance to computing or, when they have seen the relevance, they've not been able to get from the narrow ideas to what they hope aesthetics could contribute to computing.

Too many people today still assume (and some philosophers still believe) that aesthetics is principally concerned with making disinterested judgments of the quality of beauty inherent in a class of unique and autonomous objects called works of art. However, aesthetics has no unique set of objects, not only because so many "things" can be works of art, as the history of modern art has taught us, but because aesthetics is as much about people, experience, and value as it is about objects or things. And beauty is no longer a principal concern in aesthetics because it's not a principal concern in art (for a host of reasons analyzed by others elsewhere) [20]. Moreover, aesthetics is not about the fixed properties of any objects, whether works of art, natural objects, or artifacts of computing. This does not mean, intentionally or unwittingly, that aesthetics is merely subjective or that, as we sometimes hear, beauty is in the eye of the beholder. Aesthetics is not merely subjective any more than it is merely objective because beauty (understood not merely as a particular aesthetic property but as a stand-in for the entire set of aesthetic properties) is not in the subject any more than it is in the object. But where is beauty, if it is not a fixed property of any subject or object? In the language of eighteenth-century aesthetics, beauty is a *relational* property, that is, a property resulting from cognitive and affective relations or interactions among human subjects or between them and an open-ended set of works of art, natural objects, or artifacts of computing. In this light, the task of aesthetic computing is to identify, render explicit, and analyze critically the various conditions – technological, social, ontological, psychological, etc. – that make such relations or interactions possible, not just what makes them more effective, usable, communicable, pleasurable, and the like, though by understanding what makes them possible we'll presumably be in a better position to address these other concerns. Since the interactions here involve humans, and particularly since the interactions are not only between humans and objects but *among* humans (hence the need to shift from interaction to participation), aesthetic norms here are also moral and political. Again, aesthetics is able to coordinate all the dimensions of these norms better than either ethics or politics could because aesthetics has a long history of doing just that in the context of art.

On this account, aesthetics is a natural ally of computing because computing also traffics in objects lacking fixed properties, as is evident in Lev Manovich's discussion of the word "object" in the Introduction to *The Language of New Media*. Expressions such as Artificial Intelligence, Virtual Reality, Simulation, and Second Life likewise involve computer-based "realities" and objects that are not fixed. Also, in the field of scientific visualization involving, say, molecular biology, the data that are visualized are inaccessible to human senses since there is no light at the molecular level. So the data do not constitute objects in the usual sense of the word and their visualizations have no objective correlates. This means that there is no single objective way to visualize molecular data, no essential visualization of them just waiting to be discovered by a computer scientist (though any visualization is always constrained by scientific methodologies and goals). Moreover, this means these "objects" remain invisible even after they have been visualized, so it makes no sense to say that the visualization of molecular data have fixed properties (other than in the broadest sense of data properties – i.e., qua numbers and codes). Looking at this description of scientific visualization, computer scientists working in scientific (and other forms of) visualization should feel at home in aesthetics because molecular (and other) data are very similar to contemporary works of art: they too are not (necessarily) objects; they are more conceptual than sensuous, even when they assume sensuous form(s); and they are not imitations of objective realities against which they can be judged, so they can take numerous forms, subject to the limits of visualization and the methodological structures and goals of science (or art).

Now, if aesthetic computing is as much about human interactions as about objects or properties, a key question here is what makes these interactions *aesthetic*. How can we delimit the open-ended range of human-computer interactions and isolate those that are specifically aesthetic, especially if beauty is not a fixed property and is actually an effect of these interactions rather than a criterion for identifying them? [21] This question is both easier and harder to answer in the case of aesthetic computing than it might be in aesthetics more generally; easier, because the interactions have to involve some computing activities, artifacts, or the like, which, for the most part, are easier to identify than works of art, which have proven to be very elusive in recent years; yet harder too, because what is *aesthetic* about human interactions involving computers? The answer to this last question is that the norms implicitly embedded and enacted in the various layers of computing are what introduce the aesthetic dimension (hence database aesthetics, information aesthetics, etc., depending on which layer of the computing stack is involved).

The open-ended nature of aesthetic computing may create consternation among some computer scientists, or at least that has been my experience while researching, lecturing, or teaching about aesthetic computing. For there is a tendency to expect that aestheticians should provide objective norms (concepts, criteria, or the like) that can then serve as practical guides for researchers in computing (the field of "criticism" in computing sometimes embodies this tendency). If followed, however, this tendency would make aesthetics a field external to computing that is then applied to it. By contrast, I've proposed a model of aesthetic computing that operates only within computing by rendering explicit the aesthetic norms that are always already implicit and operative in the layers of the computing stack. Any new norms will have to emerge from within computing practices, just as new norms are introduced within artistic practices. In the end, aesthetics is either internal to computing or has little critical relevance to it.

### 26.16.6 Endnotes

1. J. M. Coetzee, *Diary of a Bad Year* (New York: Penguin, 2008). See also Jaron Lanier: "Information systems need to have information in order to run, but information underrepresents reality" – *You Are Not A Gadget: A Manifesto* (New York: Knopf, 2006).

2. As the Editor of the *Encyclopedia of Aesthetics* (New York: Oxford University Press, 1998), I'm not suggesting that the general should exclude the particular – it's all a matter of balance

3. Caroline A. Jones, "Introduction," in Jones, Ed. *Sensorium: Embodied Experience, Technology, and Contemporary* Art (Cambridge: MIT Press, 2006).

4. For more on participatory art, see, e.g., Claire Bishop, Editor, *Participation* (London & Cambridge: Whitechapel Gallery MIT Press, 2006); and *Artificial Hells: Participatory Art and the Politics of Spectator-*

*ship* (London: Verso, 2012). Nicolas Bourriaud, *Relational Aesthetics* (France: Les Presse Du Reel, 2002). Rudolf Freiling, Editor, *The Art of Participation: 1950 to Now* (San Francisco & London: San Francisco Museum of Modern Art and Thames & Hudson). Pablo Helguera, *Education for Socially Engaged Art* (New York: Jorge Pinto Books, 2011). Grant Kester, *Conversation Pieces: Community and Communication in Modern Art* (Berkeley: University of California Press, 2004); and *The One and the Many: Contemporary Collaborative Art in a Global Context* (Durham: Duke University Press, 2011). Nato Thompson, Editor, *Seeing Power: Art and Activism in the Age of Cultural Production* (New York: Melville House, 2012); and *Living as Form: Socially Engaged Art from 1991-2011* (Cambridge: MIT Press, 2012).

5. To give another example, Lev Manovich argues that people often point positively to the user-generated content available now online (e.g., anime music videos, political mashups) as evidence of artistic freedom or creativity on the internet (even enhanced democracy), yet they fail to reflect critically on the fact that this content follows implicitly embedded and enacted industry templates and conventions or reuses professionally produced content. Manovich, "Art After Web 2.0" in *The Art of Participation: 1950 to Now*.

6. Since I'm a philosopher, it's likely inevitable that my perspective on this new field is going to be general. But such generality is also due to the fact that aesthetics is a conceptual and normative field, though it clearly must be linked to the empirical reality of computing if it's going to have any efficacy as a mode of critical thinking that is internal to computing.

7. Mary Flanagan and Helen Nissenbaum have developed "values at play," a conception of critical play that identifies and transforms the values embedded and enacted in computer (and other) games. As I see it, their

approach is a good example of aesthetic computing because they render explicit the implicit norms of games. But they do not appeal to aesthetics. In fact, they seem to shun it, perhaps because Flanagan is an artist and seems to adopt uncritically the anti-aesthetic stance common in contemporary art, while Nissenbaum is a philosopher who doesn't yet appreciate the critical value of aesthetics. This is unfortunate, I think, because aesthetics provides exactly the kind of conceptual and critical resources Flanagan and Nissenbaum are developing as they analyze and create games that embed and enact transformative values. See Flanagan, *Critical Play: Radical Game Design* (Cambridge: MIT Press, 2009); and Flanagan and Nissenbaum, *Values at Play* (forthcoming).

8. Aesthetic Computing began at a conference in Dagstuhl, Germany, in 2002, from which emerged a manifesto published in *Leonardo* in 2003, and an anthology, *Aesthetic Computing*, Paul Fishwick, Ed. (Cambridge: MIT Press, 2006) (which I reviewed in *Leonardo On-line Reviews* (January 2007): http://www.leonardo.info/reviews/jan2007/aest_kelly.html Computational aesthetics, which is also called (or linked to) algorithmic aesthetics or exact aesthetics, has been traced back to the 1930s; see Gary Greenfield, "On the Origins of the Term 'Computational Aesthetics'"; and Florian Hoenig, "Defining Computational Aesthetics," in *Computational Aesthetics in Graphics, Visualization and Imaging*, I. Neumann, M. Sbert, B. Gooch, W. Purgathofer, Editors (2005), pp. 9-12 and 13-18. Database Aesthetics can be traced back to at least 1999; see Victoria Vesna, Editor, *Database Aesthetics* (Minneapolis: University of Minnesota Press, 2007). For examples of Digital Aesthetics, see Sean Cubitt's website: http://www.ucl.ac.uk/slade/digita/; and Johanna Drucker, *SpecLab: Digital Aesthetics and Projects in Speculative Computing* (Chicago: University of Chicago Press, 2009). Information Aesthetics has an active website: http://infosthetics.com/. See

also the SIGGRAPH Information Aesthetics Showcase in 2009: http://www.siggraph.org/s2009/galleries_experiences/information_aesthetics/ F or an example of Network Aesthetics, see Warren Sack, "Network Aesthetics," in *Database Aesthetics*, pp. 183-210. For an example of Software Aesthetics, see Stephan Diehl and Carsten Görg, "Aesthetics and the Visualization and Quality of Software," in Fishwick, *Aesthetic Computing*, pp. 230-37. There are also various websites devoted to this topic. And there's also Visual Aesthetics, discussed extensively elsewhere in this Encyclopedia.

9. Goodman, "Art in Action," in *Encyclopedia of Aesthetics*, pp. 322-25. For an account of Goodman's relevance to aesthetic computing, see John Lee, "Goodman's Aesthetics and the Language of Computing," in *Aesthetic Computing*, pp. 29-42.

10. Manovich speaks of the cultural layer in addition to the computing layer, but I'm envisioning aesthetic computing that integrates rather than separates these layers. See *The Language of New Media* (Cambridge: MIT Press, 2001).

11. As Fishwick clarifies in his Encyclopedia entry, aesthetic computing is different from computer or digital art, that is, digital technology applied to the arts. "Aesthetic computing" refers to the impact of artistic practices and aesthetic principles on the field of computing, so the influence flows *from* art and aesthetics *to* computing. For example, computer scientists are looking to learn from artists how to conduct critiques of their prototypes for new technologies (as artists do of their new works); how best to visualize data in scientific, information, or knowledge visualization; and how to understand the balance between form and function or, more typically in computing, beauty and usability in new technologies, especially those involving user interfaces. As these kinds of influence of art on computing are developed, aesthetics is a natural third party since art always involves some type of aesthetics.

12. Zadie Smith, "Generation Why?" (Review of *The Social Network,* a film directed by David Fincher, with a screenplay by Aaron Sorkin; and Jaron Lanier, *You Are Not a Gadget: A Manifesto* (New York: Knopf, 2010), in *New York Review of Books* (November 25, 2010): http://www.nybooks.com/articles/archives/2010/nov/25/generation-why/?pagination=false

13. See, e.g., Olav W. Bertelsen and Søren Pold, "Criticism as an Approach to Interface Aesthetics," NordiCHI '04, October 23-27, 2004; Lars Erik Udsen and Anker Helms Jørgensen, "The Aesthetic Turn: Unravelling Recent Aesthetic Approaches to Human-computer Interaction," *Digital Creativity*, 16, 4 (2005): 205–16; Jeffrey Bardzell, "Interaction Criticism and Aesthetics," *Proc. of CHI'09*. ACM Press (2009), 2357-66, and Jeffrey Bardzell, "Interaction Criticism: An Introduction to the Practice," *Interacting with Computers*, 23 (2011) 604–21. See also Olav W. Bertelsen: "Tertiary Artifacts at the Interface," in *Aesthetic Computing*, ed. Paul Fishwick (Cambridge: MIT Press, 2006), pp. 357-368. According to Bertelsen, "human-computer interaction requires understanding of the aesthetics of computing technology," that is, "how computing technology is experienced and 'experienceable.' Input from aesthetic computing is greatly needed in human-computer interaction" (p. 359). In explaining what he has in mind, Bertelsen analyzes the work of Marx Wartofsky, a philosopher of art and science. I think this is a very good article in aesthetic computing, even if one does not accept the Wartofsky framework, because Bertelsen clarifies aesthetics in a way that is philosophically sound, linked to art and science, and relevant to computing.

14. See, e.g., Kirsten Boehner, Rogério DePaula, Paul Dourish, and Phoebe Sengers, "Affect: From Information to Interaction," CC 05, *Proceedings of the Dicennial Conference on Critical Computing* (New York: ACM Press), pp. 59-68. See also the MIT Lab for Affective Computing: http://affect.media.mit.edu/

15. While any other discourses or disciplines implicated in this normative complex could critique its own type of normativity, only aesthetics is able to critique the normativity in all its complexity. For example, when Ken Goldberg installed "Demonstrate" (2004) in Sproul Plaza on the campus of the University of California at Berkeley, his project raised all sorts of issues and aesthetics is arguably at the center of them all. He set up a robotic webcamera for six weeks (24/7) that could be manipulated (zooming in, taking photographs, and the like) by people in remote locations, allowing somebody in Tokyo, say, to conduct surveillance on people in the Berkeley plaza. Although it was technology that made this installation possible, it clearly was not just an engineering project because of the consequences of remote surveillance on unsuspecting people in an open plaza on the campus of a public university. There were legal issues, starting with the question of the privacy rights of the people under surveillance, in particular because, as I understand it, the camera was not calibrated tightly enough at first so it was able to scan beyond the parameters intended for the project. In addition, because this project was also construed as an art work, there were also issues of artistic freedom, not only on behalf of Goldberg (and perhaps the people conducting the surveillance) but for the people in the plaza; for they were no longer as strictly constrained in their public behavior because they were participating in a work of art (apparently, some people engaged in or at least simulated sex acts under the protection of artistic freedom). Finally, the project commemorated the 40th Anniversary of the Berkeley-led Free Speech Movement, so there were important political issues at stake too because the movement was subjected to surveillance in its time, albeit without today's more sophisticated technology. Aesthetic critique is able to make sense of the normative complexity (technical, legal, ethical, political) of a project like Goldberg's *Demonstrate* because, again, aesthetics has a long history

of analyzing works of art with this same type of normative complexity.

16. Warren Sack argues, as I understand it, that the recognition of the aesthetic (as well as other nontechnical) dimensions of software and computing was evident from the early days of computing. See his website: http://people.ucsc.edu/~wsack/

17. Roger Malina, "A Forty-Year Perspective on Aesthetic Computing in the *Leonardo* Journal," in Fishwick, *Aesthetic Computing*, p. 48. The other, weak claim is that aesthetics may help computer scientists "achieve their [existing] objectives more easily, quickly, or elegantly" (p. 47).

18. For similar developments in other sciences, see, e.g., *Aesthetic Science: Connecting Minds, Brains, and Experience*, Arthur P. Shimamura and Stephen E. Palmer, Eds. (New York: Oxford University Press, 2012).

19. Fishwick, *Aesthetic Computing*, p. 13.

20. On the fate of beauty in modern art, see, e.g., Arthur C. Danto, *The Abuse of Beauty: Aesthetics and the Concept of Art* (Chicago: Open Court Press, 2003); Elizabeth Prettejohn, *Beauty and Art: 1750-2000* (New York: Oxford University Press, 2005); and Wendy Steiner, *Venus in Exile: The Rejection of Beauty in 20th Century* Art (Chicago: University of Chicago Press, 2001).

21. There a long-standing discussion of the open nature of art works in the history of contemporary aesthetics. See, e.g., Umberto Eco, *The Open Work*, tr. A. Cancogni (Cambridge: Harvard University Press, 1989; originally published in 1962).

## 26.17 COMMENTARY BY ROGER MALINA

How to cite this commentary in your report

### Roger Malina



Roger F. Malina is an art-science researcher, astronomer and editor. He is a Distinguished Professor of Arts and Technology and Professor of Physics at the University of Texas, Dallas where he is developing an Art-Science R and D and Experimental Publishing program. He is a Directeur de Recherche of the CNRS and former Director of the Observatoire Astronomique de Marseille Provence at A...

Roger Malina
Roger Malina is a member of The Interaction Design Foundation

Paul Fishwick has been formulating for a number of years his concept of "Aesthetic Computing", broadly defined as the application of the theory and practice of aesthetics to computing; in the process an "embodied" formal language is advocated. In my view this approach becomes particularly pertinent if, in the process, the methods and content of computing as science and engineering is changed and enhanced. I have called this kind of goal the 'strong case' for art-science interaction [1] where the interaction goes beyond the demonstrative or pedagogical. It is perhaps no accident that at the Dagstuhl workshop a 'manifesto' was issued [2] as there was a general feeling by the participants that the possible contributions of the arts, design and humanities to computer science were not generally accepted outside a group of enthusiasts.

A number of developments, some of which are referred to by Fishwick are mentioned, but here I would like to address a number of issues which are included in the concept of aesthetic computing but go beyond it. Fishwick talks of aesthetic computing addressing the different elements of formal languages which are number, data, model and software.

Here it is perhaps useful to add Denning's [3] seven principles of computing; these have the advantage of being process oriented and helps focus areas of possible art and design intervention:

1. Computation: What can and cannot be computed
2. Communication: Reliably moving information between places
3. Coordination: Effectively using many computers
4. Recollection: Representing, storing and retrieving information from media
5. Automation: Discovering algorithms for information processes
6. Evaluation: Predicting performance of complex systems
7. Design: Structuring software systems for reliability

This significantly broadens the landscape of possible intervention of aesthetic computing approaches, and indeed many of these areas have been barely engaged by the arts and humanities to date.

### 26.17.1 Big Data Transition and the Crisis of Representation.

A major development in the last decade is sometimes referred to as the “big data” transition [4]. As data volumes and rates continue to grow at accelerating rates scientific disciplines go through transformational changes. Astronomy was perhaps the first discipline to make this transition with the emergence of virtual observatory strategies on both data archives and software. Fields such as genomics soon followed and now all areas of business and culture are impacted (see for instance [5]). This has led to what might be called a “crisis of representation” and the emergence of new disciplines such as infoviz and dataviz. It has rapidly become apparent that the problem is no longer one of ‘communication’ of the content of data via illustration techniques (e.g. [6]) but rather the problem becomes one of immersion in data which can no longer be thought of as ‘objects’ but rather as a “fluid”; hence strategies of reification referred to by Fishwick are proving inadequate as we enter media territory. A number of researchers have been seeking to expand the reach of ‘image science’ into this new territory that requires semiotic approaches not yet developed. Most data is never analysed or viewed, and new kinds of “technologies of attention” are required to help navigate and isolate data that has particular content or meaning. This crisis of representation is a good area for the arts and humanities to be involved and will be a long term agenda for aesthetic computing; and as emphasised by Fishwick the key issue is embodiment or how data is put in forms that are apprehensible via the human senses. A number of artists have been prominent in exploratory projects for navigating through data such as Donna Cox [7] and Ruth West [8]. There have been a number of exhibitions that have sought to display the variety of approaches (e.g. siggraph information aesthetics [9]). It seems to me that Fishwick pays insufficient attention to these developments which are rapidly leading to new research areas.

### 26.17.2 Artificial Life Art, Visual Mathematics and Embodied Code

Paul Fishwick points out briefly the area of serious gaming as one area where code can be seen as being embodied. It seems to me there is a far larger area of algorithmic art, visual mathematics, artificial life art which have much stronger examples of embodiment of code. The development of algorithmic art by computer art pioneers such as Michael Noll, Roman Verostko, Harold Cohen to name just a few has already established a 50 year history of making algorithms apprehensible by human vision and hearing using aesthetic means and objectives. The area of visual mathematics (see Michele Emmer and his two Leonardo books on the visual mind [10]) has already provided success stories of how aesthetic methods have led to scientific, or mathematical, discoveries. Following the Santa Fe Institute workshops that established visibly the bases of artificial life [11], and its more recent applications to synthetic biology, the arts community rapidly picked up the challenge with the development of artificial-life art projects that has led to a proliferation of projects in robotics, virtual worlds, interactive installations and other ways of bringing code into physical contact with the human senses. Recently Leonardo Journal published [12] a selection of works as part of the 10th anniversary of the VIDA artificial life competition, the pre-eminent forum for artificial life art practitioners.

### 26.17.3 Translation as a possible method for Aesthetic Computing

In section 26.10 Fishwick articulates the elements of a method for aesthetic computing. In recent discussions with colleagues Rainer Schulte and Frank Dufour at the University of Texas, Dallas [13], I have been impressed by their work that seeks to apply the humanities based methods of translation studies to the problems posed by computer generated or mediated forms of creative "writing" (whether text, image, sound or indeed multi-medial and multi sensorial). There are issues of media "essentialism' as pointed out by Fishwick in his section 26.2.4 on Media, where each medium has specificities that allows or prevents certain concepts to be translated from one medium to another. But in addition - as pointed out by Fishwick - the limitations of technology may constrain certain form of embodiment. He gives the example of the experience of early word processing software where the writer had to ‘stop and wait’ for the microprocessor to catch up; similarly now certain microprocessor steps occur at faster rates than the cycle times of the human cortex, so that the act of translation from code to embodied perception requires the slowing down and time stretching of phenomena. The act of translation from the culture of the “formal languages of computing” to the “formal languages of the arts and literature”, or vice versa, requires methodologies from the humanities in order to create meaningful embodiment strategies for aesthetic computing.

### 26.17.4 Intimate Science

Elsewhere I have written [14] of the general problem of how to enable cultural appropriation of scientific phenomena inaccessible to the human senses. I have called this the agenda of ‘intimate science’ that many artists are now involved in. As pointed out by philosophers many of the ways that we conceive of the world are built from our experiences from birth onwards. Our ideas of causality, or more generally of explanatory systems, are fed by our interactions with the world via our bodies and senses. But science now deals with many phenomena that are not only beyond the “amplification” or “augmentation” of our senses but are inherently ‘non-commensurate’ with the way our senses operate. This problem perhaps emerged most clearly in the case of quantum mechanics where our basic ontologies are no longer applicable (objects can be both wave and particle) and concepts of causality (in the case of entanglement) totally foreign to our experience in the macro world. I would argue that we run into similar issues in the concept of emergence in complexity science. Certainly we run into similar issues in making sense of general relativity and the presence of distortions in space-time and the structure of space itself; there is no way to experience gravity waves as a human being. It seems to me that Fishwick’s agenda for aesthetic computing in a way transposes the problem of making science intimate into the field of computer science. Computers have different internal logics than human bodies, different teleologies, and only through the mechanisms of embodiment that he discusses can we begin to 'imagine' the way that a computer "imagines", The exercise of transposing 'models' into 'maquettes' is one such process of cultural appropriation.

### 26.17.5 STEM to STEAM

During the last two years the U.S National Science Foundation, in partnership with the U.S National Endowment for the Arts, have been organizing a number of workshops that bring together the research communities in Science and Engineering and the creative communities in Arts and Design [15]. These initiatives were responding to the sociological fact that there is a growing body of research practice that bridges Science and Engineering to Art and Design; in some cases School of Arts and Design find themselves engaged very similar research agendas to Science or Engineering departments with of course different, or overlapping, outcomes in mind. Sometimes unusual trans-discliplinary collaborations are involved, in other cases artists and designers find themselves in the role of inventors and technological innovators. This development was first recognized in information technology. The "Mitchell" report "Beyond Productivity" [15] in 2003 laid out the problems and opportunities. Similarly developments are now occurring in other areas of science and engineering. As a result of these workshops the NSF has put in place two research contracts to stimulate network development via the SEAD initiative [16], and a trans disciplinary documentation platform via the XSEAD contract [17]. An early outcome of these workshops was the acceleration of the concept of 'turning STEM into STEAM'. After thirty more years of national efforts in the U.S to develop the STEM workforce pipeline, the U.S is faced with shortage of trained scientists and engineers. As articulated by the STEM to STEAM movement (see for instance the congressional testimony organized by John Maeda, President of the Rhode island School of Design [18]) we need to integrate the Arts and Design, or more generally the Arts and Humanities, into Science Technology Engineering and Mathematics education and research strategies. Some of the fastest growing computer science related programs are in the computer arts, gaming, social media; and as pointed out by theorists in innovation studies, the process of social and cultural innovation is playing a stronger and stronger role in successful adoption of disruptive technologies. Expertise in visualization and image science areas that are grounded in the arts and humanities are promising areas that emerge from the aesthetic computing agenda that is outlined by Fishwick. As pointed out by Fishwick our institutions of higher education are badly organized to address these research agendas; the program that Fishwick has been developing at the University of Florida is one exemplar of possible approaches.

Fishwick addresses strategies of aesthetic computing for the formal language construct which are number, data, model and software. If we add Denning's seven principles of computing namely Computation, Communication, Coordination, Recollection, Automation, Evaluation, and Design it is clear that aesthetic computing is part of a larger ensemble of arts and humanities research strategies that offer the opportunity of making major contributions to computer science in the coming decades. As I write these comments, there is a large online discussion on "The New Aesthetics", a discussion that credits its source as James Bridle's blog "The New Aesthetics" [20] in May 2011. With a starting point that computing is now culturally integrated into our way of being in the world, the discussion (see for instance Ian Bogost [20], for a rebuttal) has been lively - indicating that we are only at the beginning of aesthetic computing.

## 26.17.6 References


1. Malina, Roger (2011), The Strong Case for Art-Science Interaction, Retrieved from http://vectors.usc.edu/thoughtmesh/publish/120.php

2. Paul Fishwick et al., 2003, 'Aesthetic Computing Manifesto'. Leonardo, 36, Issue No 4,

3. Peter Denning, http://cs.gmu.edu/cne/pjd/GP/GP-site/welcome.html

4. The fourth paradigm, Data Intensive Scientific Discovery, EDITED BY Tony Hey, Stewart Tansley,and Kristin Tolle, 2009 Microsoft Corporation, ISBN 978-0-9825442-0-4.

5. http://datajournalism.stanford.edu/

6. Edward Tufte, http://www.edwardtufte.com/tufte/

7. Donna Cox, http://www.ncsa.illinois.edu/~cox/

8. Ruth West, http://www.atlasinsilico.net/gallery.html

9. 2003 Siggraph Information Aesthetics Show case, http://www.siggraph.org/s2009/galleries_experiences/information_aesthetics/

10. Michele Emmer, The Visual Mind, Leonardo Books, 1993, http://leonardo.info/isast/leobooks/books/emmer.html

11. Christopher G Langton (1998). Artificial life: an overview. MIT Press. ISBN 0262621126 ,

12. Leonardo Journal Vida Gallery, http://www.leonardo.info/isast/journal/toc411.html

13. http://translation.utdallas.edu/

14. http://www.sonicacts.com/portal/index.php/roger-malina-intimate-science-or-artists-in-the-dark-universe/

15. Beyond Productivity : http://www.nap.edu/openbook.php?record_id=10671&page=235

16. http://sead.viz.tamu.edu/index.html

17. http://www.slideshare.net/ironman28/xsead-presentation-scalar-11912

18. http://stemtosteam.org/

19. James Bridle, http://new-aesthetic.tumblr.com/

20. Ian Bogost, http://www.wired.com/beyond_the_beyond/2012/04/ian-bogost-the-new-aesthetic-needs-to-get-weirder/

## 26.18 COMMENTARY BY SOPHIA KRZYS ACORD

How to cite this commentary in your report

### Sophia Krzys Acord

© *Sophia Krzys Acord*

Sophia Acord is Associate Director of the Center for the Humanities and the Public Sphere at the University of Florida. She holds Ph.D. and M.Res. degrees in Sociology from the University of Exeter (UK) and a B.A. from Swarthmore College

(USA). Acord's research draws on her background in the sociology of art and science and technology studies in order to explore the impact of digital tec...

Sophia Krzys Acord
Sophia Krzys Acord is a member of The Interaction Design Foundation

The hand knows what the mind cannot tell. This classic adage about embodied cognition, familiar to artists (cf: Sudnow, 1978) and craftsman (I like: Crawford, 2009), plays an important role in representing the complex ways in which we come to learn and understand the knowledge and substance at work in building our physical worlds. This adage also reveals the shortcomings of the perpetuating Cartesian superiority of mind over body, which privileges language as an authoritative mediator of knowledge. Paul Fishwick's career work to develop and formulate the novel field of Aesthetic Computing demonstrates an exciting and important way in which notions of sensory and bodily experience can be brought to bear on the most formal of languages: computer code. In this commentary, I will give a short definition of embodied cognition, a brief background on how thinking aesthetically about interaction can reveal important dimensions of knowledge-making, and then posit a few ways in which aesthetic computing may be transformational for the evolving digital humanities.

Cognition refers to the mental processes involved with gaining knowledge, including those involved in producing and understanding language. To cite an entry on "embodied cognition" from another excellent open-access scholarly encyclopedia, the Stanford Encyclopedia of Philosophy, we can see cognition as *embodied* when aspects of an individual's body "beyond the brain" play a significant constitutive role in one's ability to intake, process, and develop an understanding of new knowledge. New research in the cognitive sciences provides further support for the importance of embodiment in conceptual learning. As philosopher of mind Alva Noë (2006) described in summarizing this work, the sense of touch, not vision, should be our model for thinking about perception; we acquire new content through active inquiry and exploration. (This is not a new idea of course, Vichean philosophy — see Vico, 1725 — drew on the work of Aristotle to argue that men can only know what they make.) The understanding that knowledge is an action — something we do in concert with material objects, bodies, and environments — is also supported by much qualitative research in my home field of the sociology of the arts (cf: Acord and DeNora, 2008; Sutherland and Acord, 2007).

Seen from this vantage point, aesthetic computing is a move that fits into a broader theoretical paradigm interested in exploring the non- and quasi-cognitive aspects of behavior, knowledge-production, and interaction, as well as the important roles played by materials, technologies, and objects in the worlds we make. As MIT social scientist Sherry Turkle describes in her 2007 edited volume, *Evocative Objects*, the physical objects in our lives are anchors of our memories, thoughts, and action; how we interact with them demonstrates that thought and feeling are linked. Similarly, music sociologist Tia DeNora (2000) pointed out that aesthetic materials, like the songs we hear, are accomplices in our everyday lives; they allow us to undertake tasks that we could not accomplish without them. (Any Zumba instructor will be familiar with this power of music.) Even earlier studies in science and mathematics support this point. Looking at how shoppers in the grocery store use mathematics, Jean Lave (1988) demonstrated that cognition is an interactive process between persons acting and the settings in which their activity is constituted. This has also been discovered in a range or professions, including: design engineers (Henderson, 1999), cookie manufacturers (Streeck, 1996), and ship navigators (Hutchins, 1995). As Fishwick rightly points out, incorporating aesthetic encounters into learning software design reunites the mind and body of the computer scientist such that the physical coding (or serious gaming) experience can build understanding of more abstract analytic concepts. Creating embedded virtual experiences for learning code, or otherwise bringing real world bodily metaphors into software design, is significant in the student interaction with formal language. The resources we have at our hands with which to make meaning influence what we can know.

In this chapter, Fishwick identifies the challenge of aesthetic computing as "connecting humans with computers". In an evaluation of one of his University of Florida aesthetic computing courses, Fishwick, *et al*. (2005) discovered that many of the undergraduate students felt that aesthetic computing was time consuming, but particularly useful in explaining computing concepts to non-engineers. In digging more deeply into the theory behind aesthetic computing, however, I posit that there are additional important opportunities to connect humans and computers here to advance knowledge in computer science and also the humanities disciplines.

As sociologists of science show, producing scientific knowledge requires moving from 'dirty', 'fuzzy', and hands-on experiences to abstract and codified representations (cf: Latour and Woolgar, 1979). As a result, scientific results and findings are translations of our human experiences that may distort what it is that we really know. Ong (1982) makes a similar argument about the technology of written human languages: learning a written language entails a transformation of consciousness; we begin thinking with words, rather than speaking our thoughts. While Fishwick cites Mark Johnson and George Lakoff's good work to show that our embodied experiences are present in language through metaphor, Ong also demonstrates that written languages risk eliminating processes of embodiment by positioning words and written language notations as artificial mediators of what we know and how we can express it. How can we exit language to study language concepts?

I see aesthetic computing as offering an opportunity for computer scientists (and their students) to engage differently with software design by creating a new and embodied experience to 'play' (cf: Huizinga, 1944) with the concepts upon which formal language notation is built. By side-stepping formal language notation as a mediator of our knowledge of mathematical relations, and engaging with the body as a different kind of mediator, aesthetic computing may enable new ways of thinking about software design. Briefly, let us consider the case of art as an aesthetic activity, in which art-making or participation is a way to externalize and reflect upon 'felt' experiences in order to grasp (but also to extend) linguistically mediated situations. Art can be a place to work through alternative constructions and implementations of our understandings; in this way art may enable healing, conflict resolution, and social movements (Acord and DeNora, 2008). Similarly, as education scholar Donald Schön (1987) observes of architecture students, engaging in the hands-on 'making' and interrogation of more abstract architecture concepts through building physical models creates opportunities for the on-the-spot experimentation, problem-solving, and tinkering that he terms "reflection-in-action". Importantly, this embodied tinkering can create opportunities for questioning and altering language design concepts that may result in amendments to formal language systems. (The hand may feel something differently than how the mind classified it.) In this chapter, Fishwick observes, 'as we break open the lid of the black box containing the atomic elements of normally hidden data, formulas, code, and models, we find that computing is theatre all the way down.' If computing is theatre, the trick is to treat it as such: a place to engage in embodied play to not simply repeat well-rehearsed formal language concepts but also to tinker with or improve upon them. This potential for 'improvisation' is, to me, one of the most exciting potentials of aesthetic computing.

Finally, Fishwick describes aesthetic computing as a terrain to bring together computer scientists, artists, and humanities scholars to extend our models of meaning-making in the digital age. Earlier in the chapter, however, Fishwick draws on the work of Sherry Turkle and the example of the digital watch to discuss how deep abstract concepts in computing (as a formal language) affect the thinking of users. The idea that formal language shapes (and, by inference, limits) how we think is particularly frightening to some scholars in the humanities who would like to use computing technology to expand interpretive possibilities (not delimit them). (As Jaron Lanier, 2010, notes: we should use gadgets, not the other way around.) Johanna Drucker, in particular, has written about the tensions between formal languages based upon mathematics and their use with images and digital humanities data that question formalized and established regimes of meaning (cf: Drucker, 2001, 2009, 2011). Aesthetic computing, which enables questioning and play with formal language concepts themselves (not simply their written notation as code), may well be the solution to creating genuinely new and open-ended interpretive interfaces. As Dexter, et al. (2011) argue in a recent special issue in Culture Machine on the digital humanities, "The functional role of aesthetics plays out most richly when contextualized as part of the creative process of software development" (pp. 16-17). In other words, opportunities to transform interactive computing environments are enhanced when programmers see themselves as embodied. I posit, then, that to advance the digital humanities we must not only think about code aesthetically or symbolically, as does the research area of critical code studies, but take a step further and think differently about the mental/physical constructions upon which coding is based. In this way, aesthetic computing provides the physical and bodily tools to dramatically rethink possibilities for cultural computing in the digital age.

### 26.18.1 References


- Acord, Sophia Krzys and DeNora, Tia (2008). Culture and the arts: From art worlds to arts-in-action. *The Annals of the American Academy of Political and Social Science*, 619(1): 223-237.
- *Crawford, Matthew (2009) Shop Class as Soul Craft: An Inquiry into the Value of Work. New York: The Penguin Press.*
- DeNora, Tia (2000) *Music in Everyday Life*. Cambridge: Cambridge University Press.

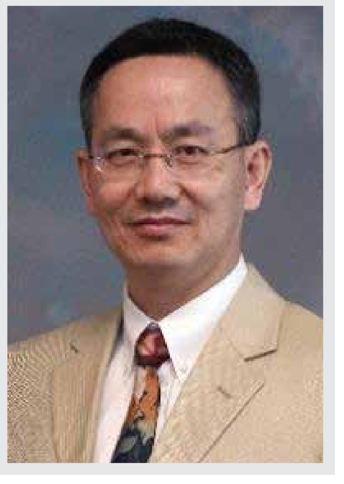
- Dexter, Scott, Melissa Dolese, Angelika Seidel, and Aaron Kozbelt (2011) On the embodied aesthetics of code. *Culture Machine*, 12.
- Drucker, Johanna (2001) Digital ontologies: The ideality of form in/and code storage: Or: Can graphesis challenge mathesis? *Leonardo*, 34(2): 141-145.
- --- *(2009) SPECLAB: Digital Aesthetics and Projects in Speculative Computing*. *Chicago, London: University of Chicago Press*.
- --- (2011) Humanities approaches to interface theory. *Culture Machine*, North America, 12.
- Fishwick, Paul, Timothy Davis, and Jane Douglas (2005) Model representation with aesthetic computing: Method and empirical study. *ACM Transactions on Modeling and Computer Simulation*, 15(3): 254—279.
- Henderson, Kathryn (1999) *On Line and On Paper: Visual Representations, Visual Culture, and Computer Graphics in Design Engineering*. Cambridge, MA: The MIT Press.
- Huizinga, Johan (1944) *Homo ludens: A study of the play element in culture*. Translated by G. Steiner, 1970. London: Paladin.
- Hutchins, Edwin (1995) *Cognition in the Wild*. Cambridge, MA: The MIT Press.
- *Lanier, Jaron (2010) You Are Not a Gadget: A Manifesto*. *New York:* Alfred A. Knopf.
- Latour, Bruno and Steve Woolgar (1979) *Laboratory Life: The Construction of Scientific Facts*. Beverly Hills: SAGE Publications.
- Lave, Jean (1988) *Cognition in Practice: Mind, Mathematics, and Culture in Everyday Life*. Cambridge: Cambridge University Press.

- Noë, Alva (2006) *Action in Perception*. Cambridge, MA: The MIT Press.
- Ong, Walter J. (1982) *Orality and Literacy: The Technologizing of the Word*. New York: Methuen.
- Schön, Donald A. (1987) *Educating the Reflective Practitioner*. San Francisco: Jossey-Bass Publishers.
- Streeck, Jürgen (1996) How to do things with things: Objects trouvés and symbolization. *Human Studies*, 19: 365-384.
- Sudnow, David (1978) *Ways of the Hand: The Organization of Improvised Conduct*. Cambridge, MA: Harvard University Press.
- Sutherland, Ian and Sophia Krzys Acord (2007). Thinking with art: From situated knowledge to experiential knowing. *The Journal of Visual Art Practice*, 6(2): 125-140.
- *Turkle, Sherry (ed.) (2007) Evocative Objects: Things We Think With. Cambridge, MA: The MIT Press.*
- *Vico, Giambattista. (1725) Scienza Nuova*. (*The First New Science*, edited and translated by Leon Pompa, Cambridge: Cambridge University Press, 2002.)

## 26.19 COMMENTARY BY KANG ZHANG

How to cite this commentary in your report

### Kang Zhang





Kang Zhang is Professor and Director of Visual Computing Lab, Department of Computer Science at the University of Texas at Dallas. He is also a Board Director of Vital Art and Science Inc., USA. He holds a B.Eng. degree in Computer Engineering from University of Electronic Science and Technology of China, a Ph.D. degree from University of Brighton, UK, and an Executive MBA degree from th...

Kang Zhang
Kang Zhang is a member of The Interaction Design Foundation

Fishwick's edited book published in 2006 had formally introduced the concept of Aesthetic Computing [1]. Based on the then definition, *aesthetic computing* addresses the question "how can the theory and techniques in the traditional visual art help beautify modern technology outputs and products and enhance their usability?". It includes the aesthetic design of computer algorithms, simulation, visualization [4], human-machine interfaces, and high-tech products, so that users are highly engaged and thus usability is enhanced. An interesting example of aesthetic computing is to apply Kandinsky's aesthetics to Java programming [3]. Malina [2] highlighted the aesthetic computing activities published in the *Leonardo* journal over the last forty years. Fishwick's new chapter on aesthetic computing takes a more specific and operational view on aesthetic computing, focusing on the notion of embodied formal language. In analogy to artists who were considered to be the first virtual reality creators, computer scientists may interpret objects dynamically via bodily simulations, possibly within virtual reality environments.

A related new discipline, but conceptually in reverse direction, is *computational aesthetics* that aims at answering the question "how can the computer automatically generate various forms of visually aesthetic expressions?" [6]. In other words, computational aesthetics investigates how the modern technology helps arts. The technology serves to create tools that can enhance the expressive power of visual art and heighten human understanding of aesthetic evaluation, perception, and meaning.

The viewpoints and concepts expressed in this chapter, some of which bear deep roots in arts, science and technology, are stimulating and of great interest to anyone who is interested in both computing (or mathematics) and art. Technology has advanced to such a level that art and design become increasingly important and relevant to science and technology, whereas the importance of our technical knowledge is decreasing. This trend will continue and the argument is indirectly supported by a recent discovery that the Internet and search technology are changing our brain and also how we think, since we no longer need to memorize, just need to know how to search to find information we need.

1. P. Fishwick (Ed.) **Aesthetic Computing**, MIT Press, Cambridge, 2006.

2. R.F. Malina, **A Forty-Year Perspective on Aesthetic Computing in the Leonardo Journal**, in: P. Fishwick (Ed.) *Aesthetic Computing*, MIT Press, 2006, 43-52.

3. C.B. Price, **From Kandinsky to Java (The Use of 20th Century Abstract Art in Learning Programming)**, *ITALICS*, Vo.6, No.4, October 2007, 35-50.

4. K. Zhang, **From Abstract Painting to Information Visualization**, *IEEE Computer Graphics and Applications*, May/June 2007, 12-16.

5. K. Zhang, **Book Review: Aesthetic Computing edited by Paul Fishwick, MIT Press, 2006, ISBN 0-262-06250-X**, *Journal of Visual Languages and Computing*, Vol.18, No.6, December 2007, Elsevier Science Inc., New York, 613-616.

6. K. Zhang, S. Harrell, and X. Ji, **Computational Aesthetics - On Complexity of Computer-Generated Paintings**, *Leonardo Journal*, MIT Press, 2012, Vol.45, No.3 (in press).

## 26.20 COMMENTARY BY DAVID J. THERRIAULT

How to cite this commentary in your report

### David J. Therriault

*© David J. Therriault*

Dr. Therriault is an Associate Professor in the School of Human Development and Organizational Studies in the College of Education at the University of Florida. He received his undergraduate degree in psychology from the University of New Hampshire and his M. A. and Ph.D. in cognitive psychology from the University of Illinois at Chicago. He was formerly a Postdoctoral Fellow at Florida ...

David J. Therriault
David J. Therriault is a member of The Interaction Design Foundation

In his chapter on Aesthetic Computing, Paul Fishwick has created a very accessible and compelling argument for applying embodiment to human-computer interfaces. Fishwick's case is built by examining the different levels of presence in embodied examples. Can you imagine yourself in the landscape of a painting, or virtually manipulating objects while doing math, or becoming a character in a textually described world? Indeed you can, and more importantly there may be learning benefits to doing so. Fishwick provides rich examples of the use of embodiment in software (e.g., a steampunk obesity machine crafted from barrels of water to explain dynamic systems) that provide us with a glimpse into the potential of aesthetic computing. In the balance of this commentary, I briefly discuss the cognitive history and my own experiences researching embodiment, and argue that Fishwick's insights provides readers with a true glimpse into the future of programing.

Traditional views of cognitive psychology assume that information processing makes use of abstract symbols. Since the 1950's, the manipulation of abstract/amodal symbols has become the cornerstone of theories examining memory, reading, and thinking. Let us take as an example the psychology of reading. Kintsch's (1974, 1998) influential research on reading comprehension was built from propositions (i.e., abstract idea units). Kintsch also provided evidence for the psychological reality of propositions, largely shaping our current understanding of what happens in our minds when we read. However, there was always a slice of representation missing from our reading experiences (often referenced as the symbol grounding problem)--a sense of presence or embodiment. How do we explain how propositions acquire meaning, how do we truly experience what we have read? In many ways, Fishwick is tackling the same symbol grounding problem (but within a programming framework).

Embodiment most simply put is attempting to understanding mind through the experiences and perceptions of our bodies (e.g., perceptual symbols, or grounded cognition). Interestingly, research examining how the body can influence our understanding has been most prominent in discourse psychology, specifically text comprehension. Researchers such as Barsalou (1999), Glenberg (1997), Lakoff and Johnson (1999), and Pecher & Zwaan (2005) have all argued that an embodied approach to cognition may have advantages over traditional views of mental representation.

My own introduction to embodiment came during the earlier 2000's working as a post doc with Rolf Zwaan. It was a wonderful experience, one were we spent the bulk of our time talking about how we would test the links between cognition and action, over good coffee or even better scotch, and see these ideas come to life in the lab. Prior to starting my position, other faculty members had cautioned me that researching grounded cognition was "kooky" and that attention on the topic would soon pass.

Since that time, there have been a host of compelling demonstrations of the psychological reality of embodiment. For example, we know that when listeners hear a story with their eyes closed they move them as if viewing the story in the real world (Spivey, Richardson, Tyler, &Young, 2000), that areas of the brain employed when doing a physical task are the same used when reading about that task (Feldman & Narayanan, 2004), and that even our judgments of morality can influence our perception, such as the perceived level of light in a room (Banerjee, Chatterjee, & Sinha, 2012). Our own research (Kaschak et al. (2002)) provided evidence that the perception of motion makes use of some of the same neural machinery needed to understand a verbal description of motion.

Research exploring an embodied view of cognition continues to flourish but there is currently no unified theory. Most researchers strongly advocate for either a symbolic or embodied view of cognition, but the field is moving away from this dichotomy. For example, Louwerse (2007) argues for exploring the relative contribution of both symbols and embodiment. Fishwick's chapter represents a truly novel approach to embodiment: applying it to create a better understanding of such things as system dynamics, number sense, or programming. This view is also refreshing, in that, embodied computing isn't intended to supplant the rich symbols tradition in coding. But embodiment may positively augment how we interact with computers in the future, and who wouldn't want that?

In sum, I find Fishwick's chapter a successful venture and one that has implications beyond undergirding (forgive the pun) the study of aesthetic computing. Programming becomes a more attractive domain for folks in my discipline to continue the study of mental representation.

### 26.20.1 References


- Banerjee, P., Chatterjee P., & Sinha, J. (2012). Is it light or dark? Recalling moral behavior changes perception of brightness. *Psychological Science*, 23 (4), 407-409.
- Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, 22, 577-660.
- Glenberg, A. M. (1997). What memory is for. *Behavioral and Brain Sciences,* 20, 1-55.
- Kaschak, M. P., Madden, C. J., Therriault, D. J., Yaxley, R. H., Aveyard, M., Blanchard, A., & Zwaan, R. A. (2005). Perception of Motion Affects Language Processing. Cognition, 94(3), B79-B89.
- Kintsch, W. (1974). *The representation of meaning in memory*. Hillsdale, New Jersey: Erlbaum
- Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. New York: Cambridge University Press
- Lakoff, G. & Johnson, M. (1999). *Philosophy in the flesh: The embodied mind and its challenge to Western thought*. New York: Basic Books.
- Louwerse, M. M. (2007). Symbolic or embodied representations: A case for symbol interdependency. In T. Landauer, D. McNamara, S. Dennis, & W. Kinsch (Eds.). Handbook of latent semantic analysis (pp. 107-120). Mahwah, NJ: Erlbaum.
- Pecher, D. & Zwaan, R. A. (Eds.) (2005). *Grounding Cognition: The role of perception and action in memory, language, and thinking*. New York: Cambridge University Press.

- Spivey, M. J., Richardson, D. C., Tyler, M. J., & Young, E. E. (2000). Eye movements during comprehension of spoken scene descriptions. *Proceedings of the Twenty-second Annual Meeting of the Cognitive Science Society*, (pp. 487-492), Erlbaum: Mawhah, NJ.

## 26.21 REFERENCES

*Anderson, Kenneth M., Taylor, Richard N. and Whitehead, E. James (2000): Chimera: hypermedia for heterogeneous software development environments. In ACM Transactions on Information Systems, 18 (3) pp. 211-245*

*Audi, Robert (ed.) (1999): The Cambridge Dictionary of Philosophy. Cambridge University Press*

*Barsalou, Lawrence W. (2010): Grounded Cognition: Past, Present, and Future. In Topics in Cognitive Science, 2 (4)*

*Beck, Dennis, Fishwick, Paul A., Kamhawi, Rasha, Coffey, Amy Jo and Henderson, Julie (2011): Synthesizing Presence: A Multidisciplinary Review of the Literature. In Journal of Virtual Worlds Research, 3 (3)*

*Bimber, Oliver and Raskar, Ramesh (2005): Spatial Augmented Reality: Merging Real and Virtual Worlds. A K Peters, Ltd*

*Biocca, Frank (1997): The Cyborg's Dilemma: Progressive Embodiment in Virtual Environments. In Journal of Computer-Mediated Communication, 3 (2)*

*Bolter, Jay David and Grusin, Richard (2000): Remediation: Understanding New Media. The MIT Press*

*Bowman, Doug A., Kruijff, Ernst, LaViola, Joseph J. and Poupyrev, Ivan (2004): 3D User Interfaces: Theory and Practice. Addison-Wesley Professional*

*Brown, John Seely, Collins, Allan and Duguid, Paul (1989): Situated Cognition and the Culture of Learning. In Educational Researcher, 18 (1)*

*Burke, Joseph (1943): Hogarth and Reynolds: A contrast in English art theory. Oxford University Press*

*Buxton, Bill (2008): The Role of the Artist in the Laboratory. In: (ed.). "Meisterwerke der Computer Kunst". pp. 29-32*

*Card, Stuart K., Mackinlay, Jock D. and Shneiderman, Ben (eds.) (1999): Readings in Information Visualization: Using Vision to Think. Academic Press*

*Cavallaro, Dani (2001): Cyberpunk & Cyberculture: Science Fiction and the Work of William Gibson. The Athlone Press*

*Cockton, Gilbert (2012). Usability Evaluation. Retrieved 6 June 2013 from [URL to be defined - in press]*

*Dede, Chris (2005): Planning for Neomillennial Learning Styles Neomillennial Learning Styles and Mediated Immersion. In EDUCAUSE Quarterly, 28 (1) pp. 18-21*

*Devlin, Keith (2006): The Useful and Reliable Illusion of Reality in Mathematics. In: Towards a New Epistemology of Mathematics. Workshop at GAP.6 Freie Universität Berlin 2006.*

*Dewey, John (1934): Art as Experience. Perigee Trade*

*Diehl, Stephan (2007): Software Visualization: Visualizing the Structure, Behaviour, and Evolution of Software. Springer*

*Dieterle, Edward, Dede, Chris and Schrier, Karen (2007): "Neomillennial" Learning Styles Propagated by Wireless Handheld Devices. In: Lytras, Miltiadis D. and Naeve, Ambjorn (eds.). "Ubiquitous and Pervasive Knowledge and Learning Management: Semantics, Social Networking and New Media to Their Full Potential". pp. 35-66*

*Diffey, T. J. (1995): A note on the meanings of the term 'aesthetic'. In The British Journal of Aesthetics, 35 (1) pp. 61-66*

*Dourish, Paul (2001): Where the Action Is: The Foundations of Embodied Interaction. MIT Press*

*Eades, Peter and Zhang, Kang (1996): Software Visualization. Series on Software Engineering and Knowledge Engineering. World Scientific Pub Co Inc*

*Feldman, Jerome (2006): From Molecule to Metaphor: A Neural Theory of Language. A Bradford Book*

*Feldman, Jerome and Narayanan, Srinivas (2004): Embodied meaning in a neural theory of language. In Brain and Language, 89 (2) pp. 385-392*

*Fishwick, Paul A. (2012). Aesthetic Computing Class. Retrieved 1 January 2012 from University of Florida: http://www.cise.ufl.edu/~fishwick/ac/2012*

*Fishwick, Paul A. (ed.) (2006): Aesthetic Computing. The MIT Press*

*Fishwick, Paul A. (2007a): Customized Visual Computing: The Aesthetic Computing Method. In: Ferri, Fernando (ed.). "Visual Languages for Interactive Computing: Definitions and Formalizations (Premier Reference Source)". pp. 425-435*

*Fishwick, Paul A. (ed.) (2007b): Handbook of Dynamic System Modeling. Chapman and Hall - CRC*

*Fishwick, Paul A. (2002): Aesthetic Programming: Crafting Personalized Software. In Leonardo, 35 (4) pp. 383-390*

*Fishwick, Paul A. (1995): Simulation Model Design and Execution: Building Digital Worlds. Prentice Hall*

*Fishwick, Paul A. (2008): Software Aesthetics: From Text and Diagrams to Interactive Spaces Paul A. Fishwick. In International Journal of Arts and Technology, 1 (1) pp. 1-13*

*Fishwick, Paul A. (2012b): A Decade of Digital Arts and Sciences at the University of Florida. In Leonardo, pp. 211-216*

*Fishwick, Paul A. and Bertelsen, Olav W. (2002). Dagstuhl Seminar Report Number 348. http://www.dagstuhl.de/Reports/02/02291.pdf*

*Fishwick, Paul A. and Park, Yuna A. (2008): A 3D Environment for Exploring Algebraic Structure and Behavior. In: Ferdig, Richard E. (ed.). "Handbook of Research on Effective Electronic Gaming in Education Set of 3". Information Science Referencepp. 546-559*

*Fishwick, Paul A., Coffey, Amy J., Kamhawi, Rasha and Henderson, Julie (2005): An experimental design and preliminary results for a cultural training system simulation. In: Simulation Conference WSC, Proceedings of the 2010 Winter 2005. pp. 799-810*

*Fishwick, Paul A., Diehl, Stephan, Prophet, Jane and Lowgren, Jonas (2005b): Perspectives on Aesthetic Computing. In Leonardo, 38 (2) pp. 133-141*

*Fuller, Matthew (ed.) (2008): Software Studies: A Lexicon (Leonardo Book Series). The MIT Press*

*Gershenfeld, Neil, Krikorian, Raffi and Cohen, Danny (2004): The Internet of things. In Scientific American, 291 (4) pp. 76-81*

*Ghezzi, Carlo and Jazayeri, Mehdi (1997): Programming Language Concepts. Wiley*

*Goldin, Gerald A and Kaput, James J (1996): A Joint Perspective on the Idea of Representa-*

*tion in Learning and Doing Mathematics. In: Steffe, Leslie P., Nesher, Pearla, Cobb, Paul, Goldin, Gerard A. and Greer, Brian (eds.). "Theories of Mathematical Learning". Routledge- pp. 397-431*

*Grau, Oliver (2004): Virtual Art: From Illusion to Immersion (Leonardo Books). The MIT Press*

*Greenfield, Adam (2006): Everyware: The Dawning Age of Ubiquitous Computing. New Riders Publishing*

*Grossman, Lev (2009). Steampunk: Reclaiming Tech for the Masses. Retrieved 19 April 2012 from Times Magazine: http://www.time.com/time/magazine/article/0,9171,1945343,00.html*

*Guhe, Markus, Smaill, Alan and Pease, Alison (2009): A formal cognitive model of mathematical metaphors. In: Proceedings of the 32nd annual German conference on Advances in artificial intelligence 2009. pp. 323-330*

*Hadamard, Jacques (1996): The Mathematician's Mind. Princeton University Press*

*Hassenzahl, Marc (2011). User Experience and Experience Design. Retrieved 6 June 2013 from [URL to be defined - in press]*

*Hopcroft, John E., Motwani, Rajeev and Ullman, Jeffrey D. (2000): Introduction to Automata Theory, Languages, and Computation (2nd Edition). Addison Wesley*

*Huff, Kenneth A. (2006): Visually Encoding Numbers using Prime Factors. In: Fishwick, Paul A. (ed.). "Aesthetic Computing (Leonardo Books)". The MIT Press*

*Ifrah, Georges (2002): The Universal History of Computing: From the Abacus to the Quantum Computer. Wiley*

*Ikei, Yasushi and Ota, Hirofumi (2008): Spatial Electronic Mnemonics for Augmentation of Human Memory. In 2008 IEEE Virtual Reality Conference, 53 (14) pp. 217-224*

*Johnson, Mark (2007): The Meaning of the Body: Aesthetics of Human Understanding. University of Chicago Press*

*Johnson, Steven A. (1997): Interface Culture. Basic Books*

*Johnson, Mark (1987): Body in the Mind: The Bodily Basis of Meaning, Imagination and Reason. University of Chicago Press*

*Johnson, Eric and Adamo-Villani, Nicoletta (2010): A Study of the Effects of Immersion on Short-term Spatial Memory. In Virtual Reality, 71 pp. 1-44*

Kallay, William (2011): *The Making of Tron: How Tron Changed Visual Effects and Disney Forever*. William Kallay

Kelly, Michael (ed.) (1998): *Encyclopedia of Aesthetics*. Oxford University Press, USA

Kelly, Michael, Vesna, Victoria, Fishwick, Paul A., Moere, Andrew Vande and Huff, Kenneth (2009): The state of aesthetic computing or info-aesthetics: curated panel discussion. In: *ACM SIGGRAPH 2009 Art Gallery* 2009. pp. 43-

Kivy, Peter (ed.) (2004): *The Blackwell Guide to Aesthetics*. Wiley-Blackwell

Knuth, Donald E. (1992): *Literate Programming*. Center for the Study of Language and Information

Lakoff, George and Johnson, Mark (2003): *Metaphors We Live By, 2nd edition*. University of Chicago Press

Lakoff, George and Johnson, Mark (1983): *Metaphors We Live by*. New York, Basic Books

Lakoff, George and Johnson, Mark (1999): *Philosophy in the Flesh: The Embodied Mind and Its Challenge to Western Thought*. New York, Basic Books

Lakoff, George and Nunez, Rafael (2001): *Where Mathematics Comes From: How the Embodied Mind Brings Mathematics into Being*. Basic Books

Laurel, Brenda K. (1991): *Computers as Theatre*. Reading, MA, Addison-Wesley Publishing

Lave, Jean and Wenger, Etienne (1991): *Situated Learning: Legitimate Peripheral Participation (Learning in Doing: Social, Cognitive and Computational Perspectives)*. Cambridge University Press

Lessig, Lawrence (2008): *Remix: Making Art and Commerce Thrive in the Hybrid Economy*. Penguin Press HC, The

Leymarie, Frederic Fol (2006): Aesthetic Computing and Shape. In: Fishwick, Paul A. (ed.). "Aesthetic Computing". The MIT Presspp. 258-288

Lowgren, Jonas (2008). Interaction Design - brief intro. Retrieved 6 June 2013 from [URL to be defined - in press]

Malina, Roger (2011). The Strong Case for Art-Science Interaction. Retrieved 19 March 2012 from thoughtmesh: http://vectors.usc.edu/thoughtmesh/publish/120.php

Manovich, Lev (2001): *The Language of New Media*. The MIT Press

*Mar, Raymond A. and Oatley, Keith (2008): The Function of Fiction is the Abstraction and Simulation of Social Experience. In Perspectives on Psychological Science, 3 (3) pp. 173-192*

*Marino, Mark C. (2006). Critical Code Studies. Retrieved 18 April 2012 from Electronic Book Review: http://www.electronicbookreview.com/thread/electropoetics/codology*

*McLuhan, Marshall (1964): Understanding Media: The Extensions of Man. The MIT Press*

*Mohr, Manfred (2011). Réflexions Sur Une Ésthetique Programmée. Retrieved 18 March 2012 from Bitforms Gallery, New York: http://www.bitforms.com/press-releases/405-2011-manfred-mohr.html*

*Nake, Frieder (2009): The semiotic engine: notes on the history of algorithmic images in Europe. In Art Journal, pp. 76-89*

*Negroponte, Nicholas (1996): Being Digital. Vintage*

*Norman, Donald A. (2004): Emotional Design: Why We Love (Or Hate) Everyday Things. Basic Books*

*Papert, Seymour (1980): Mindstorms: Children, Computers and Powerful Ideas. New York, Basic Books*

*Parsons, Thomas D and Rizzo, Albert A (2008): Initial validation of a virtual environment for assessment of memory functioning: virtual reality cognitive performance assessment test. In CyberPsychology and Behavior, 11 (1) pp. 17-25*

*Petre, Marian and Blackwell, Alan (1999): Mental Imagery in Program Design and Visual Programming. In International Journal of Human-Computer Studies, 51 (1) pp. 7-30*

*Piaget, Jean (1950): The Psychology of Intelligence. Routledge*

*Pine, B. Joseph (1999): The Experience Economy: Work Is Theater & Every Business a Stage. Harvard Business Press*

*Pine, B. Joseph and Davis, Stan (eds.) (1999): Mass Customization: The New Frontier in Business Competition. Harvard Business School Press*

*Reas, Casey and Fry, Ben (eds.) (2007): Processing: A Programming Handbook for Visual Designers and Artists. The MIT Press*

*Ritchie, Andrew C. (1968): English Painters, Hogarth to Constable. Ayer Co Pub*

*Roth, Tina, Aiken, Peter and Hobbs, Scarlette (1994): Hypermedia Support for Software Development: A Retrospective Assessment. In Hypermedia, 6 (3) pp. 149-173*

Rotman, Brian (2000): Mathematics as Sign: Writing, Imagining, Counting (Writing Science). Stanford University Press

Ryder, William (2009): A System Dynamics View of the Phillips Machine. In Proceedings of the 27th International Conference of the System Dynamics Society,

Salomon, Gavriel (1990): Cognitive Effects With and Of Computer Technology. In Communication Research, 17 (1) pp. 26-44

Samuel P. Harn Museum of Art (2011). Art in Engineering. Museum Nights at the Harn Museum of Art. Retrieved 19 April 2012 from Samuel P. Harn Museum of Art: http://www.harn.ufl.edu/octmuseumnight.pdf

Scruton, Roger (2011): Beauty: A Very Short Introduction. Oxford University Press

Sfard, Anna (1994): Reification as the birth of metaphor. In For the Learning of Mathematics, 14 (1) pp. 44-55

Sfard, Anna and Thompson, Patrick W. (1994): Problems of reification: Representations and mathematical objects. In Proceedings of the Annual Meeting of the International Group for the Psychology of Mathematics Education North America Plenary Sessions, 1

Sherman, William R. and Craig, Alan B. (2002): Understanding Virtual Reality: Interface, Application, and Design (The Morgan Kaufmann Series in Computer Graphics). Morgan Kaufmann

Slingerland, Edward (2008): What Science Offers the Humanities: Integrating Body and Culture (New Approaches to European His). Cambridge University Press

Speer, Nicole K., Zacks, Jeffrey M. and Reynolds, Jeremy R. (2007): Human brain activity time-locked to narrative event boundaries. In Psychological Science, 18 (5) pp. 449-455

Stasko, John T., Domingue, John B., Brown, Marc H. and Price, Blaine A. (eds.) (1998): Software Visualization. The MIT Press

Swade, Doron (2000): The Phillips Machine and the history of computing. In: Leeson, Robert (ed.). "A. W. H. Phillips: Collected Works in Contemporary Perspective". Cambridge University Presspp. 120-126

Thompson, Patrick W. and Sfard, Anna (1994): Problems of reification: Representation and Mathematical objects. In: Proceedings of the Annual Meeting of the International Group for the Psychology of Mathematics Education - North America 1994, Baton Rouge, USA.

Tractinsky, Noam, Katz, A. S. and Ikar, D. (2000): What is Beautiful is Usable. In Interacting with Computers, 13 (2) pp. 127-145

Tufte, *Edward R. (1983): The Visual Display of Quantitative Information. Cheshire, CT, Graphics Press*

Turkle, *Sherry (2004): How Computers Change the Way We Think. In Higher Education, 50 (21)*

Varela, *Francisco J.,* Thompson, *Evan and* Rosch, *Eleanor (1991): The Embodied Mind: Cognitive Science and Human Experience. Cambridge, Massachusetts, MIT Press*

Viégas, *Fernanda B.,* Wattenberg, *Martin,* Ham, *Frank van,* Kriss, *Jesse and* McKeon, *Matt (2007): ManyEyes: a Site for Visualization at Internet Scale. In IEEE Transactions on Visualization and Computer Graphics, 13 (6) pp. 1121-1128*

Ward, *Matthew O.,* Grinstein, *Georges and* Keim, *Daniel (2010): Interactive Data Visualization: Foundations, Techniques, and Applications. A K Peters Ltd*

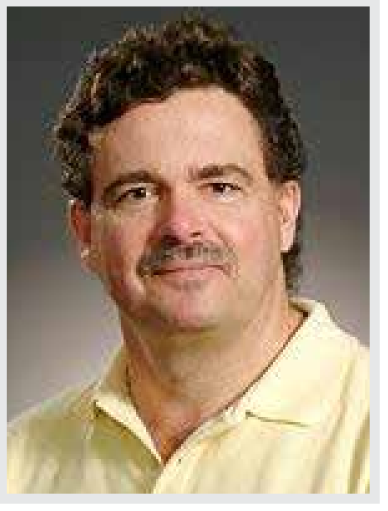
Wing, *Jeannette M. (2006): Computational thinking. In Communications of the ACM, 49 (3) pp. 33-35*

Yates, *Frances A. (1966): The Art of Memory. University of Chicago Press*

Zhang, *Kang (2007): Visual Languages and Applications. Springer*

# About the author

## Paul A. Fishwick



*© Paul A. Fishwick*

Has also published under the name of:
"Paul Fishwick"

Personal Homepage:
cise.ufl.edu/~fishwick/Welcome.html

Current place of employment:
University of Florida

Paul Fishwick is Professor of Computer and Information Science and Engineering at the University of Florida, Director of Digital Arts & Sciences Programs, and CEO of Metaphorz, LLC. He has six years of industrial experience at Newport News Shipbuilding & Dry Dock Co (now: Northrup Grumman). and at NASA Langley Research Center in Virginia. He obtained the PhD from the University of

Pennsylvania, is a Fellow of the Society for Computer Simulation International, and is highly active in modeling and simulation research, with emphasis on virtual environments, scientific modeling and workflow environments, GPU-based simulation acceleration., social network acquisition interfaces, and model-based training environments for culture and medicine. Commonly used technologies include immersive visualization, portable devices, 3D animation, and mixed/augmented reality.

## Your notes and thoughts on chapter 26

Record your notes and thoughts on this chapter. If you want to share these thoughts with others online, go to the bottom of the page at:
http://www.interaction-design.org/encyclopedia/aesthetic_computing.html

**NOTES**

# Chapter 27

# CSCW - Computer Supported Cooperative Work

*by Jonathan Grudin and Steven Poltrock.*

Computer Supported Cooperative Work (CSCW) is a community of behavioral researchers and system builders at the intersection of collaborative behaviors and technology. The collaboration can involve a few individuals or a team, it can be within or between organizations, or it can involve an online community that spans the globe. CSCW addresses how different technologies facilitate, impair, or simply change collaborative activities.

The CSCW community revolves around a journal and two conference series, one typically held in North America and one in Europe. Books and academic courses followed, and relevant papers appear in other conferences as well. Pointers to these resources conclude this chapter.

## 27.1 THE EMERGENCE OF CSCW

In 1984 Irene Greif and Paul Cashman coined the acronym CSCW for an invited workshop focused on understanding and supporting collaboration. Technology capable of supporting a group of people was so expensive that workplace deployment was the sole focus. A major topic was email, which in 1984 was poorly designed, not interoperable across different platforms, and used primarily by researchers. The first open CSCW conference was held in 1986. CSCW soon became the principle research forum for the collaboration that was newly enabled by emerging client-server PC and workstation networks. Despite the severe processing and memory constraints in those early days, these networks created new possibilities.

What inspired these researchers? In 1988, Greif published *Computer-Supported Cooperative Work: A book of readings* (Greif 1988). Four of the first five papers describe the inspirational research led by Douglas Engelbart between 1963 and 1984. Engelbart is best known for inventing the mouse, but he had a far broader vision for augmenting human intellect and building high-performance teams through technology. The NLS system he and his colleagues developed and used (See Figure 27.1 A-B) included many features that took decades to become widely used, including desktop and video conferencing.



Figure 27.1 A-B: Douglas Engelbart and staff using NLS to support 1967 meeting with sponsors - probably the first computer-supported conference. The facility was rigged for a meeting with representatives of the ARC's research sponsors NASA, Air Force, and ARPA. A U-shaped table accommodated setup CRT displays positioned at the right height and angle. Each participant had a mouse for pointing. Engelbart could display his hypermedia agenda and briefing materials, as well as the documents in his laboratory's knowledge base.

Another source of inspiration was more recent behavioral research (also included in Greif's book), much of it centered on minicomputer office automation systems. These systems supported *computer-mediated communication*, a term still in use, and the technologies were often called *groupware*, reflecting the early focus on small groups or teams. Use of this label declined as organization-wide deployment became common and as collaboration features were integrated into more applications.

In theory, CSCW could cover any aspect of cooperative work in which digital technology plays a role. In practice, the CSCW research field reflects the interests of its participants. For example, by the mid-1980s database systems were already a maturing technology used in many organizations and not covered within CSCW. Research centered on communication, such as use of email and videoconferencing prototypes, and on small-group interaction, such as collaborative text editing and drawing. Over time, technological advances and shifting interests of CSCW researchers broadened the scope of CSCW. It came to cover interaction within units of all sizes, using both fixed and mobile technologies. Social media are now a major focus.

The terms *computer, support, cooperative,* and *work* have all been transcended. CSCW encompasses collaboration that uses technologies we do not call computers, collaboration in which technology plays a central rather than a support role, uses that involve conflict, competition, or coercion rather than cooperation, and studies of entertainment and play.

A strong European branch of CSCW formed with a somewhat different focus. Several papers in the 1988 conference from Nordic countries described participatory or cooperative design approaches. In 1989 the European conference series began, with strong German and British participation. Liam Bannon and Kjeld Schmidt (Bannon and Schmidt 1989) outlined a vision, some of which came to pass and some of which did not. Whereas much North American involvement was initially from commercial software developers and telecommunications companies, Europe drew mainly from government and academic research focused on large enterprises that at that time typically designed and developed software in-house.

The European approach was also more firmly grounded in theory, with activity theory a particularly strong influence (for example, see Engeström and Middleton, 1998).

## 27.2 CSCW TECHNOLOGY

The framework in Table 27.1 is a useful way to conceptualize collaboration technologies and their similarities and differences. Human behaviors that contribute to collaboration may be roughly divided into three categories: communication, sharing information, and coordination. People may engage in these behaviors at the same time (real time collaboration) or at different times (asynchronous collaboration). Technologies or technology features have been developed to support each of these six components of collaborative behavior.

| | **Real time** | **Asynchronous** |
|---|---|---|
| **Communication** | • Telephone<br>• Video conferencing<br>• Instant messaging<br>• Texting | • Email<br>• Voice mail<br>• Blogs<br>• Social networking sites |
| **Information sharing** | • Whiteboards<br>• Application sharing<br>• Meeting facilitation<br>• Virtual worlds | • Document repositories<br>• Wikis<br>• Web sites<br>• Team workspaces |

| | | |
|---|---|---|
| **Coordination** | • Floor control<br>• Session management<br>• Location tracking | • Workflow management<br>• CASE tools<br>• Project management<br>• Calendar scheduling |

Table 27.1: A two-dimensional collaboration framework with examples of technology features or products found within each cell.

Communication tools remains central to CSCW, even as email is studied less and microblogging more. Communication via voice, video, text conferencing, instant messaging, and text messaging have been explored. Waves of research into prototype desktop video systems appeared in the late 1980s, mid-1990s, and early 2000s. As video communication finally blossoms, past CSCW studies covering a range of complex social and interface issues will likely contribute (Poltrock and Grudin, 2005). Social networking sites such as Twitter and Facebook (see Figure 27.2 A-B) blend communication and information sharing features.

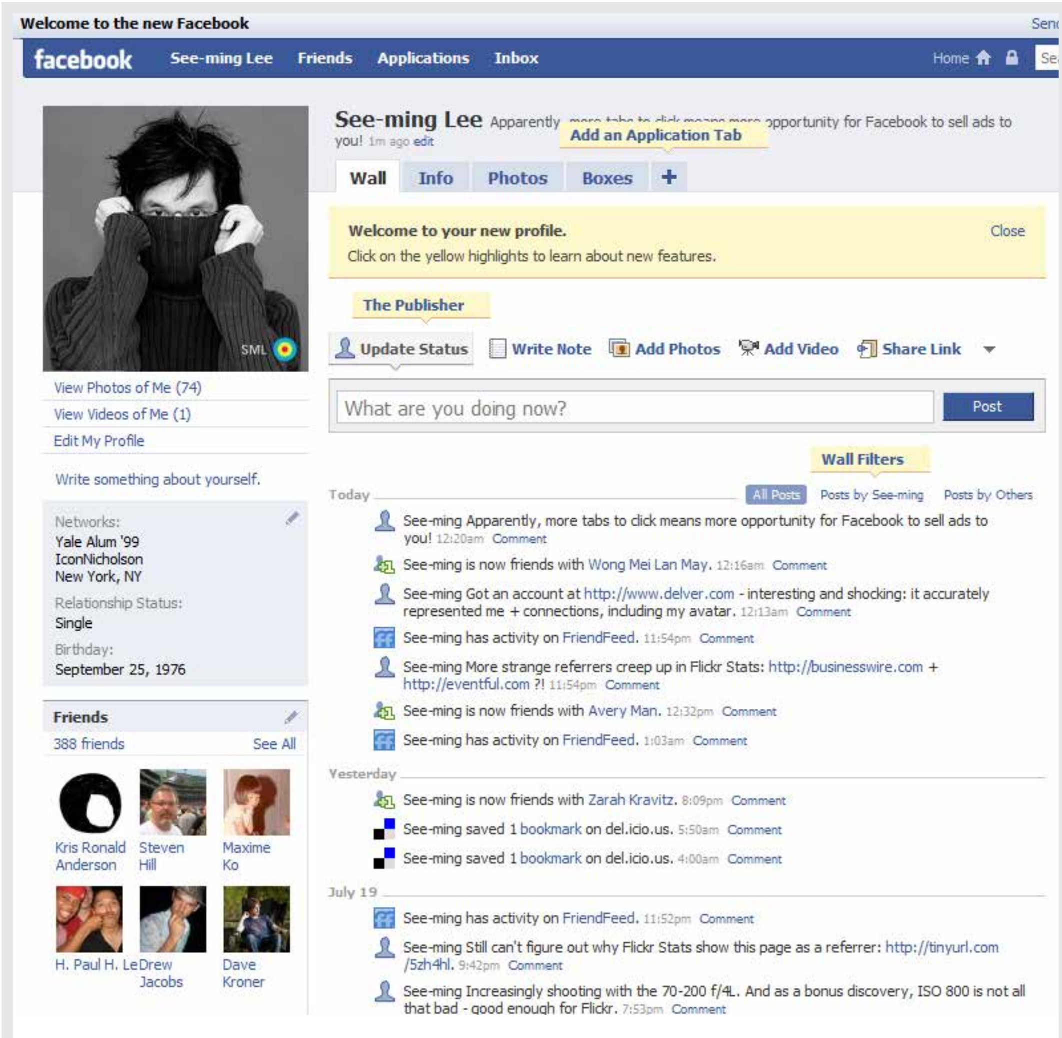



Figure 27.2: A user profile in the Facebook social networking site as it looked in 2010.



Information repositories provide a way to share information. Early studies focused on document management systems, more recently attention has shifted to wikis and Wikipedia, the mountain of freely-accessible information including a complete edit history that is swarmed over by an army of graduate students who analyze it in diverse ways. An early influential study combined visualization and analysis to examine conflict visualized through the history of edit changes (Viégas et al, 2004). Other topics include information reliability, contributor reliability, incentive systems, image contribution, and Wikipedia administration. Team workspaces such as Microsoft's SharePoint or Google Wave (see Figure 27.3) provide a managed repository for a team's artifacts and tools for communicating and sharing information with one another.

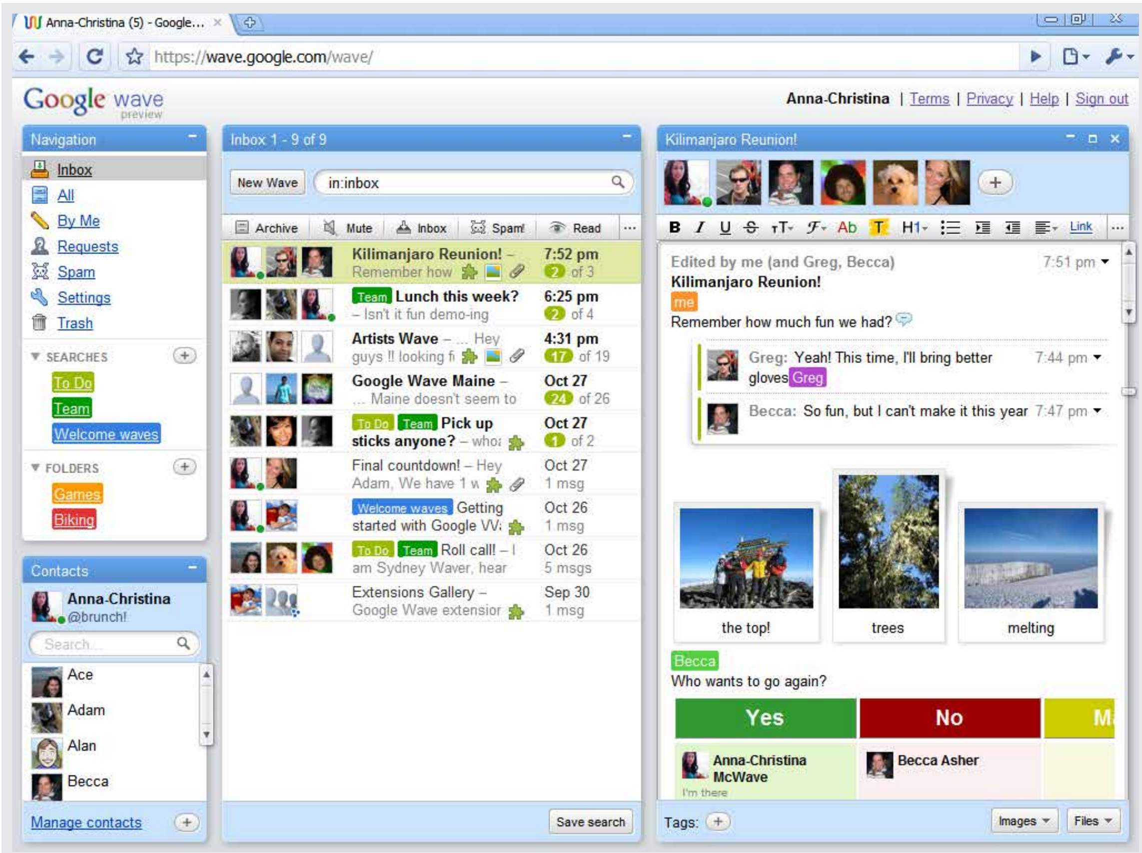





FIGURE 27.3 A-B: Google Wave conversation and collaboration. In 2009, Google started beta testing Google Wave, a real-time collaboration environment that Google hoped would eventually displace email and instant messaging. However, Google announced in August 2010 that it had decided to stop developing Wave as a standalone project, due to insufficient user adoption.

Coordination technologies employed in the workplace such as meeting support systems, group calendars, workflow management systems, and computer-aided software engineering systems were an early focus of CSCW. They gave way to studies of how people coordinate in the absence of (or despite) coordination management technologies. For example, Bowers et al (1995) studied the problems that deployment of workflow technology created in a large printing enterprise. Social networking also enables a new generation of coordination technologies whether mobile and location-aware real-time (e.g., Foursquare and Google Latitude - Figure 27.4 - or asynchronous - e.g., Groupon's coordination of purchasing decisions).


FIGURE 27.4: Google Latitude (initial release February 5, 2009) shows your friends on a map--as long as they've agreed to share their location.

## 27.3 SUPPORTING GROUPS, ORGANIZATIONS, AND COMMUNITIES

Another lens for considering CSCW research is the social unit: small groups, teams, projects, organizations, and communities. Many distinctions arise on the continuum from dyads to globe-spanning communities. Three research and development clusters have been (i) the social psychology of groups or teams and technologies to support them; (ii) organizational behavior and support; and (iii) community analysis and support.

### 27.3.1 Social psychology and group support

Social psychologists seek general principles of social behavior that are independent of organizational context. Participants in controlled experiments are often students who are assigned to work in groups. The psychologists' hope is to generalize the results of controlled experiments to the more variable conditions of the workplace environment. These generalizations may be questionable, but findings of these experiments may suggest behaviors that should be carefully examined in the workplace.

Kraut (2003) discusses why the research approach of social psychology was of limited value to technology developers. This research approach led, however, to Joseph McGrath's (McGrath 1991) invaluable framework that characterizes team behavior in terms of three functions (production, group well being, and member support) and four modes (inception, problem solving, conflict resolution, and execution), as shown in Table 27.2 below.

Easily overlooked

| | Production | Group well-being | Member support |
|---|---|---|---|
| **Inception** | Production demand and opportunity | Interaction demand and opportunity | Inclusion demand and opportunity |
| **Problem-solving** | Technical problem solving | Role network definition | Position and status achievements |
| **Conflict resolution** | Policy resolution | Power and payoff distribution | Contribution and payoff distribution |
| **Execution** | Performance | Interaction | Participation |

Principle focus of attention

Table 27.2: McGrath's (1991) framework for categorizing team behaviors.

The key to understanding the framework's utility is to focus on the columns. Organizations are obsessed with demonstrating that a new technology or process yields a "return on investment," measured as increased performance: the lower left cell, the production function and the execution mode. This apparently logical goal has two significant drawbacks: It is often impossible to prove that a communication or collaboration tool yields positive performance effects in real-world settings, so much time and money is squandered in futility. Lab studies of technology use overwhelmingly focus on impacts in the lower left cell. Second, with a laser focus on performance, it is easy to overlook that positive or negative impacts in other cells can have crucial indirect consequences.

For example, no one could prove a productivity benefit for email. Eventually people stopped questioning it. On the other hand, group support systems (electronic meeting rooms, a major focus of research in the 1980s and commercialization attempts in the 1990s) did well in controlled studies but were never commercially successful. Why? An analysis by Dennis and Reinicke (2004) attributes this to the lack of support for group well-being and member support. One participant in a meeting conducted using a group support system told us that it was the most unpleasant meeting he had experienced in his life, despite its success at accomplishing its stated objective.

Some technologies that show no positive effects in lab studies that focus on performance can provide benefits in other cells. They can aid in conflict resolution or problem-solving, enable people to achieve recognition or status, and so on. Videoconferencing can have subtle effects that are difficult to measure in terms of return on investment: It can assist conflict-resolution or problem-solving (Williams, 1997), and if people like it, it could strengthen group ties.

### 27.3.2 Organizational behavior and support

As CSCW formed in the mid-1980s, much of the research into group and team behaviors was shifting from social psychology to organizational psychology. Both groups initially participated in CSCW, but organizational psychologists interested in technology use soon shifted to alternative publication outlets. Little of their work appears in the CSCW literature, but it has important implications.

Organizations are made up of groups and teams, of course, but the challenge of supporting them is heightened by the diversity of organizational contexts. People in different parts of an organization often respond very differently to the same application (Perin, 1991; Grudin, 2004). Mintzberg (1984) identifies the major parts of an organization shown in Figure 27.5: the executives (strategic apex), managers (middle line), individual contributors (operating core), the people formulating work processes (technostructure), and the support staff. People in each of these five areas work and use technology differently.

The central three organizational parts have an interesting relationship to the three rows in Table 27.1 in this chapter: communication, information sharing, and coordination. Individual contributors often communicate extensively and informally with one another. They also have few meetings, cannot delegate work, and their work activity is visible to others in the organization. For managers, a high priority is sharing structured information in documents, spreadsheets, and slide decks. They also have many meetings, can delegate some work, and face a trade-off between the sensitivity of some tasks and the efficiency gained by informing others of their status. Executives coordinate activities of different groups. They are also constantly in meetings, delegate tasks, and their meeting schedules are sensitive. Improving performance, the production-execution cell in McGrath's framework, requires different technologies and patterns of use in different parts of an organization. For any technology used by all employees, Mintzberg's groups differ in the features that they appreciate and those that they dislike.

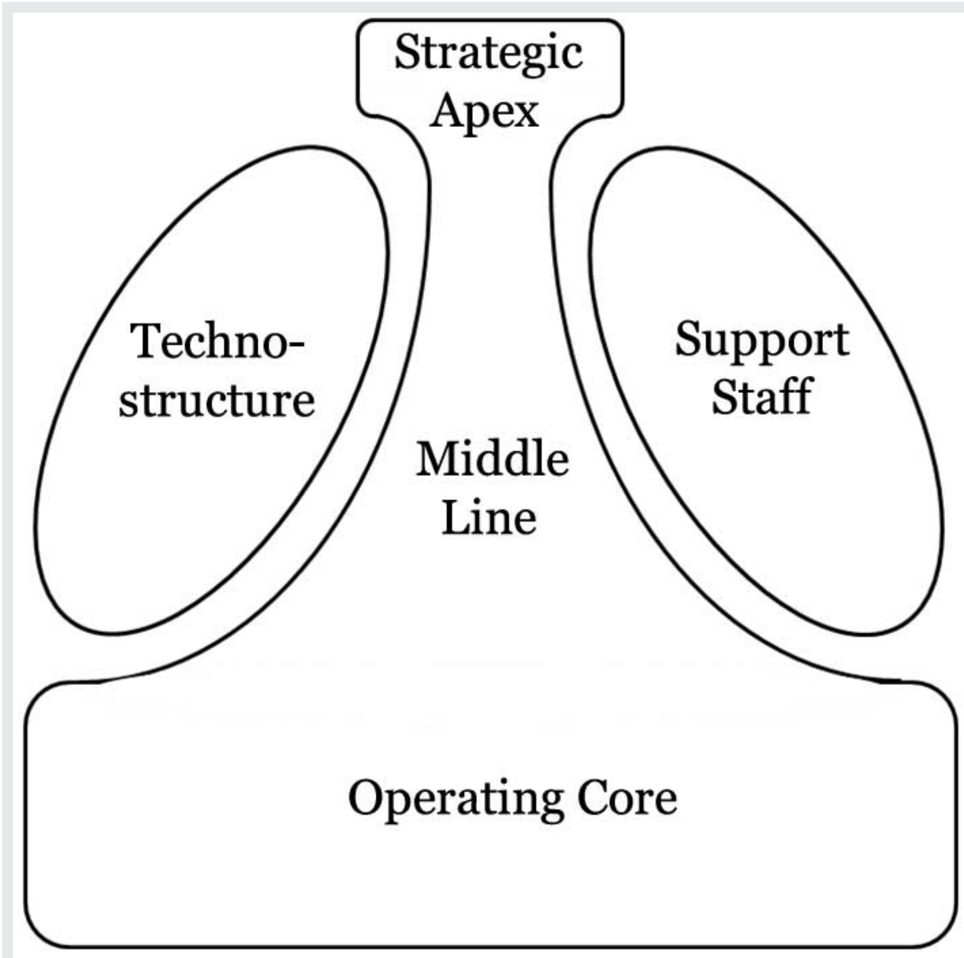



Figure 27.5: A graphical summary of Mintzberg's (1984) model of key organizational parts.



### 27.3.3 Ethnography forms a bridge

In the 1980s, North American CSCW sought commercial applications to sell to the multitude of small groups in the world and European CSCW focused on enterprise systems. As the 1990s progressed, application developers discovered that they could not ignore the organizational context of groups, and enterprises began acquiring and adapting commercial applications. As the two CSCW camps came together, ethnography was positioned to assist both.

Lucy Suchman's group at Xerox PARC spearheaded the influx of ethnographers into CSCW. Qualitative research methods revealed the effects of context on social behaviors, provided deeper understanding of the means of collaboration, and offered new insights into how technology could support or disrupt collaboration.

Luff et al (2000) is a compendium of studies and analyses that show the potential for ethnographic approaches to inform system design.

### 27.3.4 Networks and community support

While researchers and developers focused on group and organizational support, use of the Internet and then the Web grew, producing the bubble that burst spectacularly at the turn of the century. Warren Buffet observed that when the tide goes out, we discover who was swimming without trunks. As it happens, some were swimming with trunks—ecommerce and web-based communities took hold. Network technologies provided support for existing communities and created many new ones.

Early CSCW research indicated that evangelizing and facilitation are needed to achieve necessary levels of group and organizational participation in brainstorming and other collective activity. On the Internet, with enough participants, a critical mass of participation is more readily found. Crowdsourcing and viral spread of use appeared. Collaborative filtering was one of the most influential research topics (see Herlocker, Konstan & Riedl, 2000, and work cited therein).

Large scale of use also gives rise to "big data" and a path for obtaining insight into how and when people communicate and share information. Network analysis, data mining, and machine learning gained prominence. Wikipedia's transparency has led to scores of analyses. Search engines and proprietary tools such as Facebook restrict access, but studies surface.

## 27.4 CHALLENGES AND OPPORTUNITIES

Stepping back, we see steady progress in the use of collaboration technologies. Up close, it often proceeds more slowly than expected. Introducing a new technology is initially disruptive; it requires a leap of faith to believe that eventually greater efficiency or effectiveness will result. In addition, we are social beings, with preferred ways of interacting represented in our genes and in social conventions developed over a long time. We are not always consciously aware of them, but a technology that violates them is unlikely to succeed.

### 27.4.1 Design and Adoption Challenges

The following table, Table 27.3 is adapted from Grudin (1994a). The fact that it and related earlier work is among the most cited topics in CSCW is an indication that these are stubborn challenges. They affect not only developers, but also those involved in acquiring and deploying collaboration support tools.

| | |
|---|---|
| 1. | Disparities in effort required and benefits for individuals |
| 2. | Limits of informed intuition: Managers & designers beware |
| 3. | Achieving Critical Mass and avoiding Prisoner's Dilemma |
| 4. | Avoiding other social & motivational pitfalls |
| 5. | Exception-handling: The bane of workflow & other systems |
| 6. | Designing for low-frequency events |
| 7. | The difficulty of evaluation |
| 8. | Designing with an adoption process in mind |

Table 27.3: Eight design and adoption challenges.

1. When a technology requires that some of its users perform additional work without a compensating benefit, adoption may not follow.

2. A corollary is that a decision about which technology to research, develop, or deploy is generally based on the intuitions of managers who may not appreciate how the technology will be received by other group members.

3. A collaboration tool may not achieve the necessary critical mass of users. For example, one of the most actively researched CSCW applications was desktop videoconferencing; generally only some people had video, but everyone could be reached by telephone. (The converse problem is the tragedy of the commons: a technology may cease to be useful if too many people embrace it.) More insidious is the prisoner's dilemma: If everyone looks to their own best interest, perhaps by free-riding on the efforts of others, then outcomes deteriorate and use may collapse altogether.

4. Adverse social and motivational effects can result when a new tool disrupts existing channels, creates uncertainty about where to find information, and challenges existing authority structures. For example, executives who adopt a communication tool that enables them to interact directly may find that their administrators, now out of the loop, are less effective at managing their schedules and anticipating events.

5. A major contribution from ethnographic studies is the observation that group behavior often focuses on handling exceptions or unanticipated events. Technologies designed to coordinate work activities often incorporate and enforce models of standard work processes and cannot gracefully handle deviations from such models.

6. Collaborative features often must paradoxically be unobtrusive, because they are often used less frequently than other features, yet readily accessible when needed.

7. In addition, these technologies can be very difficult to evaluate…

8. …and adopt. Designers must consider what will be needed to promote successful adoption from the outset.

Mark and Poltrock (2004) studied a success case, the rapid adoption of data conferencing in a large organization, with minimal technical support and no management mandate. A system of overlapping distributed social groups overcame differences in local customs, values, and infrastructures when members of a group supported adoption by other members, regardless of their location.

### 27.4.2 Opportunities

Rapid technology change has been a driving force in this field, most clearly seen in the fact that 25 years ago, few people were on the Internet, the Web did not exist, and the computers of the time could generally do only one task at a time. The social media of today have predecessors in chat systems of decades past, but using a simple chat application then required all the resources of an expensive desktop computer. Today, feature rich applications run in the background or one window, or on an inexpensive mobile phone. The experience is constantly being radically reinvented.

Technology change has shaped not only communication and collaboration possibilities, but the relevant social issues. A nice illustration is our awareness of the activities of distant collaborators. For many years, people only saw what collaborators sent or told them. Passive awareness was technically difficult. Then, between 1992 and 1995, as local and wide area networks spread, the first three CSCW papers with 'awareness' in the title were published. The World Wide Web took hold. 'Awareness' appeared in 12 paper titles from 1996 through 1999.

The stance toward remote awareness rapidly evolved. The first paper, "Awareness and Coordination in Shared Workspaces" (Dourish and Bellotti, 1992) celebrated its achievement and potential utility. Several years later, the focus shifted to risks of *too much* awareness, as in "Techniques for Addressing Fundamental Privacy and Disruption Tradeoffs in Awareness Support Systems" (Hudson and Smith, 1996).

A similar progression occurred with another awareness tool, desktop videoconferencing. In 1992, prototype builders argued for surreptitious monitoring of colleagues: "One-way connections have advantages we are unwilling to give up. Glances allow us to maintain our awareness of colleagues without actually engaging in interaction with them... Video provides an excellent means to gain awareness unobtrusively; enforcing symmetry for the sake of privacy would undermine this functionality" (Gaver et al., 1992). Slowly, the initial tolerance of privacy invasion and the stress on maximizing technology use faded. The first step toward greater symmetry in awareness was an audible notification that one was being watched but no indication of who it was, and eventually invitation and reciprocity became the norm.

Along with the increase in capability, declining technology prices enabled researchers to get ahead of the curve. Collaboratories have been developed to support large-scale multisite efforts, primarily in scientific research, engineering, and education (see Olson and Olson, 2012). Their lofty goals enable many of them to get ample funding, but the communication and collaboration problems they encounter will be common in other settings, so studying them to glean best practices and technology requirements is a logical research strategy.

## 27.5 FUTURE DIRECTIONS

The future of the design and use of collaboration technologies is incredibly exciting. There is unlimited space for design and study, whether or not it appears under the label 'CSCW.'

Moore's law has not been revoked, and a massive industry employing a lot of ingenious people has a stake in insuring that smaller, more powerful, less expensive technologies continue to surface. Smaller sensors and effectors, larger displays, better networking, and powerful visualizations will drive innovation and open design opportunities.

Interest in multi-user simulations and virtual worlds has waxed and waned. Second Life (see Figure 27.6) attracted attention in the late 2000s. World of Warcraft is a gaming success. Dramatic high-end uses of technology that have appeared in theme parks are strong indications that compelling, fully immersive digital environments are possible and will eventually be affordably priced.

Figure 27.6: A group meeting in Second Life - circa 2010.

*Courtesy of Hildekd. Copyright: pd (Public Domain (information that is common property and contains no original authorship)).*

Opportunities on the behavioral side come from two directions. As technology is used to support more of our activities in ever finer granularity, changing those activities as it does, there is a tremendous need for — and benefit in — understanding the changes around us, some of them broad and others domain-specific. This is a golden opportunity for ethnography; organizations and societies have never evolved as rapidly and profoundly as they will in the years ahead. At the same time, the information streaming over networks presents unparalleled opportunities for data mining and information visualization, and for machine learning to find and exploit patterns in the data.

In the past, large-scale statistical analysis and ethnography were largely distinct undertakings, but in the future the two will work together, the former finding correlations and patterns, the latter enabling us to discover what they mean and identify other patterns to look for.

Education, medicine, and software development are domains in which communication, information sharing, and coordination are of wide interest. Of these, education is coming to the fore, because a key facet is information that can be readily digitized and shared. With the availability of so much information of variable quality, a new set of skills is required of students — browsing, skimming, assessing, and synthesizing to a greater degree than before.

The single largest issue that we are likely to contend with, as individuals, organizations, and societies for a very long time, is how to deal with transparency. Once information is digital it can show up anywhere at any time in the future, and much more is represented digitally every minute of every day. In this way, we learn that people do not behave as we thought they did or believe they should, people do not follow policies, regulations, or laws as closely as we imagined, social conventions are not honored consistently, and violations of all of these are prosecuted unevenly. We see chaos, inconsistency, and fallibility that was always present but not revealed. How will we react? Formulate more nuanced rules, enforce them more strictly, or become more tolerant of deviation?

The field of CSCW that had come together around groups in organizations in the 1990s has now split again, with North America highly focused on social media and online community and European CSCW focused on organizational issues and domain-specific research and development. CSCW has become more prominent in Asia and South America as well, with elements of each focus. It is plausible that these efforts will converge again as social media move into widespread use in a broad range of organizations.

## 27.6 WHERE TO LEARN MORE

*Review articles and books.*

*Ackerman, Mark S., Halverson, Christine A., Erickson, Thomas and Kellogg, Wendy A. (eds.) (2010): Resources, Co-Evolution and Artifacts: Theory in CSCW. Springer*

*Baecker, Ronald M. (1992): Readings in Groupware and Computer-Supported Cooperative Work: Assisting Human-Human Collaboration (Interactive Technologies). Morgan Kaufmann*

*Greif, Irene (ed.) (1988): Computer-Supported Cooperative Work: A Book of Readings. San Mateo, California, Morgan Kaufmann*

*Grudin, Jonathan and Poltrock, Steven (2012): Taxonomy and theory in computer-supported cooperative work. In: Kozlowski, Steve W. J. (ed.). "The Oxford Handbook of Organizational Psychology, Two-Volume Set (Oxford Library of Psychology)". Oxford University Press, USA*

*Kraut, Robert E. (2003): Applying social psychological theory to the problems of group work. In: Carroll, John M. (ed.). "HCI Models, Theories, and Frameworks: Toward a Multidisciplinary Science (Interactive Technologies)". Morgan Kaufmann*

*Olson, Gary M. and Olson, Judith S. (2012): Collaboration technologies. In: Jacko, Julie A. (ed.). "Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies, and Emerging Applications, Third Edition (Human Factors and Ergonomics)". CRC Press*

### 27.6.1 CSCW history

Grudin, Jonathan (2012): A moving target: The evolution of human-computer interaction. In: Jacko, Julie A. (ed.). "Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies, and Emerging Applications, Third Edition (Human Factors and Ergonomics)". CRC Press

*Grudin, Jonathan (1994): Computer-Supported Cooperative Work: History and Focus. In IEEE Computer, 27 (5) pp. 19-26*

*Grudin, Jonathan and Poltrock, Steven (2003): Computer-supported cooperative work and groupware. In: Zelkowitz, Marvin (ed.). "Advances in Computers". Academic Press*

## 27.6.2 Relevant Conference Series

### ECSCW - European Conference on Computer Supported Cooperative Work

2009 2007 2003 2003 2001 2001 1999 1997
1995 1993 1991 1989

Next conference is coming up 21 Sep 2013 in Paphos, Cyprus

### CSCW - Conference On Computer-Supported Cooperative Work

2012 2012 2012 2011 2010 2008 2006 2004
2004 2002 2000 1998 1996 1994 1992 1990
1988 1986

- **CollabTech** International Conference on Collaboration Technologies, held annually in Asia
  Collabtech 2011: http://www.collabtech.org/
- **IEEE CollaborateCom** International Conference on Collaborative Computing
  CollaborateCom 2011: http://www.collaboratecom.org/
- **Collaborative Technologies and Systems**
  CTS 2011: http://cts2011.cisedu.info/

General conferences with papers relevant to CSCW include CHI (ACM) and INTERACT (IFIPS). Also relevant are specialized conferences such as WikiSym (proceedings archived by ACM), ICWSM (AAAI International Conference on Weblogs and Social Media, Ubicomp, Pervasive, PDC (Participatory Design), COOP (Design of Cooperative Systems), and GROUP (ACM).

### 27.6.3 Journals and book series

**Computer Supported Cooperative Work: The Journal of Collaborative Computing** is published by Springer. http://www.springer.com/computer/journal/10606 Springer also publishes a CSCW book series: http://www.springer.com/computer/series/2861

### 27.6.4 Web sites

Wikipedia research can be found, not surprisingly, on Wikipedia:
http://en.wikipedia.org/wiki/Wikipedia:Academic_studies_of_Wikipedia
http://en.wikipedia.org/wiki/Academic_studies_about_Wikipedia.

## 27.7 REFERENCES

*EC-CSCW 89 - Proceedings of the First European Conference on Computer-Supported Cooperative Work 13-15 September, 1989, Gatwick, London.*

*Ackerman, Mark S., Halverson, Christine A., Erickson, Thomas and Kellogg, Wendy A. (eds.) (2010): Resources, Co-Evolution and Artifacts: Theory in CSCW. Springer*

*Baecker, Ronald M. (1992): Readings in Groupware and Computer-Supported Cooperative Work: Assisting Human-Human Collaboration (Interactive Technologies). Morgan Kaufmann*

*Bannon, Liam and Schmidt, Kjeld (1989): CSCW: Four characters in search of a context. In: EC-CSCW 89 - Proceedings of the First European Conference on Computer-Supported Cooperative Work 13-15 September, 1989, Gatwick, London. pp. 358-372*

*Bowers, John, Button, Graham and Sharrock, Wes (1995): Workflow from Within and Without: Technology and Cooperative Work on the Print Industry Shopfloor. In: Marmolin, Hans, Sundblad, Yngve and Schmidt, Kjeld (eds.) ECSCW 95 - Proceedings of the Fourth*

*European Conference on Computer-Supported Cooperative Work 11-15 September, 1995, Stockholm, Sweden. pp. 51-66*

Dennis, Alan R. and Reinicke, Bryan (2004): *Beta Versus VHS and the Acceptance of Electronic Brainstorming Technology. In MIS Quarterly, 28 (1) pp. 1-20*

Dourish, Paul and Bellotti, Victoria (1992): *Awareness and Coordination in Shared Workspaces. In: Proceedings of the 1992 ACM conference on Computer-supported cooperative work November 01 - 04, 1992, Toronto, Ontario, Canada. pp. 107-114*

Engeström, Yrjo and Middleton, David (eds.) (1998): *Cognition and Communication at Work. Cambridge University Press*

Gaver, William W., Moran, Thomas P., MacLean, Allan, Lovstrand, Lennart, Dourish, Paul, Carter, Kathleen and Buxton, Bill (1992): *Realizing a Video Environment: EuroPARC's RAVE System. In: Bauersfeld, Penny, Bennett, John and Lynch, Gene (eds.) Proceedings of the ACM CHI 92 Human Factors in Computing Systems Conference June 3-7, 1992, Monterey, California. pp. 27-35*

Greif, Irene (ed.) (1988b): *Computer-Supported Cooperative Work: A Book of Readings. San Mateo, California, Morgan Kaufmann*

Greif, Irene (ed.) *Proceedings of the 1988 ACM conference on Computer-supported cooperative work September 26 - 28, 1988, Portland, Oregon, United States.*

Grudin, Jonathan (2012): *A moving target: The evolution of human-computer interaction. In: Jacko, Julie A. (ed.). "Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies, and Emerging Applications, Third Edition (Human Factors and Ergonomics)". CRC Press*

Grudin, Jonathan (1994b): *Computer-Supported Cooperative Work: History and Focus. In IEEE Computer, 27 (5) pp. 19-26*

Grudin, Jonathan (1994a): *Groupware and Social Dynamics: Eight Challenges for Developers. In Communications of the ACM, 37 (1) pp. 92-105*

Grudin, Jonathan (2004): *Managerial Use and Emerging Norms: Effects of Activity Patterns on Software Design and Deployment. In: HICSS 2004 2004.*

Grudin, Jonathan (1988): *Why CSCW Applications Fail: Problems in the Design and Evaluation of Organizational Interfaces. In: Greif, Irene (ed.) Proceedings of the 1988 ACM conference on Computer-supported cooperative work September 26 - 28, 1988, Portland, Oregon, United States. pp. 85-93*

Grudin, Jonathan and Poltrock, Steven (2003): Computer-supported cooperative work and groupware. In: Zelkowitz, Marvin (ed.). "Advances in Computers". Academic Press

Grudin, Jonathan and Poltrock, Steven (2012): Taxonomy and theory in computer-supported cooperative work. In: Kozlowski, Steve W. J. (ed.). "The Oxford Handbook of Organizational Psychology, Two-Volume Set (Oxford Library of Psychology)". Oxford University Press, USA

Herlocker, Jonathan L., Konstan, Joseph A. and Riedl, John (2000): Explaining Collaborative Filtering Recommendations. In: Kellogg, Wendy A. and Whittaker, Steve (eds.) Proceedings of the 2000 ACM conference on Computer supported cooperative work 2000, Philadelphia, Pennsylvania, United States. pp. 241-250

Hudson, Scott E. and Smith, Ian (1996): Techniques for Addressing Fundamental Privacy and Disruption Tradeoffs in Awareness Support Systems. In: Olson, Gary M., Olson, Judith S. and Ackerman, Mark S. (eds.) Proceedings of the 1996 ACM conference on Computer supported cooperative work November 16 - 20, 1996, Boston, Massachusetts, United States. pp. 248-257

Krasner, Herb and Greif, Irene (eds.) Proceedings of the 1986 ACM conference on Computer-supported cooperative work December 03 - 05, 1986, Austin, Texas.

Kraut, Robert E. (2003): Applying social psychological theory to the problems of group work. In: Carroll, John M. (ed.). "HCI Models, Theories, and Frameworks: Toward a Multidisciplinary Science (Interactive Technologies)". Morgan Kaufmann

Luff, Paul, Hindmarsh, Jon and Heath, Christian (eds.) (2000): Workplace Studies: Recovering Work Practice and Informing System Design. Cambridge University Press

Mark, Gloria and Poltrock, Steven (2004): Groupware adoption in a distributed organization: transporting and transforming technology through social worlds. In Information and Organization, 14 (4) pp. 297-327

McGrath, Joseph E. (1991): Time, Interaction, and Performance (TIP): A Theory of Groups. In Small Group Research, 22 (2) pp. 147-174

Mintzberg, Henry (1984): A typology of organizational structure. In: Miller, Danny (ed.). "Organizations: A Quantum View". Prentice Hall College Divpp. 68-86

Olson, Gary M. and Olson, Judith S. (2012): Collaboration technologies. In: Jacko, Julie A. (ed.). "Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies, and Emerging Applications, Third Edition (Human Factors and Ergonomics)". CRC Press

*Perin, Constance (1991): Electronic Social Fields in Bureaucracies. In Communications of the ACM, 34 (12) pp. 74-82*

*Poltrock, Steven and Grudin, Jonathan (2005): Videoconferencing: Recent Experiments and Reassessment. In: HICSS 2005 - 38th Hawaii International Conference on System Sciences 3-6 January, 2005, Big Island, HI, USA.*

*Suchman, Lucy A. (1983): Office Procedure as Practical Action: Models of Work and System Design. In ACM Transactions on Information Systems, 1 (4) pp. 320-328*

*Viégas, Fernanda B., Wattenberg, Martin and Dave, Kushal (2004): Studying cooperation and conflict between authors with history flow visualizations. In: Dykstra-Erickson, Elizabeth and Tscheligi, Manfred (eds.) Proceedings of ACM CHI 2004 Conference on Human Factors in Computing Systems April 24-29, 2004, Vienna, Austria. pp. 575-582*

*Williams, Gayna (1997): Task Conflict and Language Differences: Opportunities for Videoconferencing. In: Hughes, John F., Prinz, Wolfgang and Schmidt, Kjeld (eds.) Proceedings of the Fifth European Conference on Computer-Supported Cooperative Work 7-11 September, 1997, Lancaster, UK. pp. 97-108*

## About the authors

### Jonathan Grudin

Has also published under the name of:
"J. Grudin"

Personal Homepage:
research.microsoft.com/en-us/um/people/jgrudin/

Current place of employment:
Microsoft Research

I work in the Adaptive Systems and Interaction Group at Microsoft Research, part of the Microsoft Corporation. My research is in human-computer interaction and computer supported cooperative work, with a particular focus on the design, adoption and use of group support technologies. Some of the work below was done in the Collaborative and Multimedia Systems Group. Prior to joining Microsoft Research, I was Professor of Information and Computer Science at University of California, Irvine. I have taught at Aarhus University, Keio University, and the University of Oslo, and worked at the MRC Applied Psychology Unit, Wang Laboratories, and MCC since earning my Ph.D. at UC San Diego.

## Steven Poltrock

© Steven Poltrock

Has also published under the name of:
"Steven E. Poltrock"

Personal Homepage:
projects.ischool.washington.edu/cir/poltrock.htm

I retired from The Boeing Company in 2009 where I was a Technical Fellow in the Research & Technology division. My research was in computer support for cooperative work, focusing on industrial applications of collaboration technology, including support for teamwork, workflow management, and knowledge management. My educational background includes degrees in engineering, mathematics, and psychology, and my career has combined my interests in all three areas. After completing a Ph.D. in psychology at the University of Washington I was a professor of psychology at the University of Denver and worked at Bell Laboratories and MCC before joining Boeing Research & Technology.

## Your notes and thoughts on chapter 27

Record your notes and thoughts on this chapter. If you want to share these thoughts with others online, go to the bottom of the page at: http://www.interaction-design.org/encyclopedia/cscw_computer_supported_cooperative_work.html

**NOTES:**

# Chapter 28

## Phenomenology

*by Shaun Gallagher.*

Phenomenology begins as a rigorous and systematic study of consciousness in the work of Edmund Husserl at the start of the 20th century. Phenomenological research has direct relevance to design since the artifacts, tools, and technologies that we make affect the way that we experience our surroundings, and this is what phenomenology studies. Emphasizing the important role that embodiment plays in perception and cognition, it investigates, among other things, affective, aesthetic, and action-oriented experience as it is informed by environmental factors and by actual and potential bodily movement. Phenomenology explores the ways that our physical and social environments, including the things and instruments in such environments matter for experience, cognition, problem solving, and for shaping our intersubjective and social interactions.

Figure 28.1: Edmund Husserl.



In Husserl's view phenomenology is a project aimed at describing the basic structures of consciousness, that is, those features of consciousness that shape the way that various objects in the world appear to us. Towards the end of the 1920s Martin Heidegger, who had worked with Husserl, offered his own somewhat different characterization of phenomenology. For Heidegger the aim of phenomenology was to provide a basic ontological analysis of human existence as 'being-in-the-world'. An important part of his analysis showed that our primary way of existing involves a pragmatic action-oriented way of being related to our environment. A third development in phenomenology can be seen in the emphasis placed on embodiment in the work of the French philosopher Maurice Merleau-Ponty. Merleau-Ponty pursued several insights offered by Husserl and furthered the analysis of embodied perception by integrating psychology and neuroscience with phenomenology and laying the groundwork for some contemporary approaches to embodied cognition. By no means do these three aspects or developments in phenomenology capture the entire scope in this field, but they do signal three areas that are directly relevant to design. Accordingly, I'll organize my account of phenomenology by focusing on these areas (See, Gallagher 2012a for a more extensive account).

## 28.1 CONSCIOUSNESS IS ABOUT SOMETHING

It is obvious that the design of objects in the built environment will have some aesthetic effects on the way we experience the surrounding world. Such aesthetic effects may range from vague and perhaps mostly unnoticed feelings that are difficult to express, to a clearly delineated pleasure or pain tied to the way that a certain object or configuration of objects appears. For Husserl this experience can be described as involving a variety of factors.

One of the first phenomenological principles is that all consciousness is characterized by intentionality, which simply means that one's consciousness is directed at something, or is about something. Perceptual consciousness, for example, always has an intentional object. Consciousness is always consciousness-of-something. When I see, I see something, some state of affairs, event, or entity. The intentional structure of consciousness can be analyzed into two aspects, which Husserl calls, referencing Greek philosophical terms, the noesis and the noema.

The noesis includes the particular character of the mental act, that is, whether the conscious act is perceptual, or is a case of memory, judgment, desire, etc., or involves some combination of these. I can see the apple on the table in front of me; or I can remember it being there; or I can judge that the apple will taste good, or I can desire to eat the apple, etc. In each case the intentional object, the apple, remains the same, but my consciousness is different. The noesis can be complex. I may simultaneously see, judge and desire, and each element in this complex act may modulate the other. The way I see something may lead me to judge it in a certain way, and this may lead me to desire it or not. It may also be the case that my desiring of something makes me see it in a certain way. In some cases my complex noetic state may be fulfilled when I pick up the apple and take a bite; or it may be disappointed if the apple doesn't taste right. Likewise, I may imagine that something has a certain appearance, and that intuition may be fulfilled (or unfulfilled) by my subsequent perceptual experience of it. Various cognitive (noetic) processes may shape my perception or vice versa. One can thus develop a rich description on this noetic side of consciousness.

The way the object appears to conscious perception (judgment, memory, etc.) is what Husserl calls the noema. This is the 'something' as it appears in my experience. I see the apple, for example, from a certain angle. The apple is presented in a certain profile. I can pick it up and manipulate it in my hand, turning it, seeing one side and then another. All of these manipulations do not change the apple itself; rather, they result in noematic modulations. There is a noematic nucleus — in each case I am seeing the apple, and that has a certain meaning or sense; but the appearances of the apple change as I manipulate it. The full noema includes these actual changes and any other possible changes that might happen, for example, if I took a bite of the apple. Changes in what I am tasting, how the apple feels to touch, and the visual appearance of the apple all enter into perceptual noematic modulations. The noema may also be conceptual, if, for example, I am making a judgment about something.

Husserl further distinguishes between the inner and outer horizons of the object — all such aspects showing up in noematic variations. The inner horizon of the object includes the way the object can change appearance, for example, as I manipulate it, or move around it, or as the light hits it differently. The outer horizon includes its relations to other things in the nearby environment, those things that constitute the spatial (or Gestalt) background of the object, its spatial context.

All such noematic aspects are in correlation with noetic aspects. Certain noematic correlates appear only when I taste as opposed to see something; or only when I see something from a certain angle and in a certain profile. Moreover, although much of what I have just said pertains to perception, it also pertains to other noetic act qualities. For example, if I am thinking about (but not perceiving) something in order to make a judgment about it, what counts as its inner or outer horizon may in fact be certain conceptual rather than perceptual relations. Thus in thinking about an apple and judging its aesthetic appeal, I may be comparing it to an orange or a pear or a lemon. More abstractly, when thinking of the concept of justice, I may be doing so by contrasting a conservative/aristocratic concept with a democratic/liberal concept of justice. Here, the various inner or outer noematic horizons of the concept of justice are conceptual.

## 28.2 CONSCIOUSNESS IS PHENOMENAL

Such considerations are part of an intentional analysis of consciousness. Consciousness also involves a certain affective feel. In contemporary discussions in the field of consciousness studies this is sometimes referred to as the phenomenal aspect of consciousness or the 'what it is like' aspect. There is not only an intentional difference, but also a phenomenal difference between tasting an apple and tasting a lemon. One has a different aesthetic experience in these two tastes. In the same way the color of an environment can be experienced as warm or cold; the surfaces as sharp or soft and what comes along with (and in) such experiences is a certain phenomenal feel — perhaps a feeling of comfort or discomfort. In his work Thing and Space, Husserl also develops the idea that our perceptual experiences of environmental objects have a kinaesthetic value. The artifacts around elicit certain kinds of movements by us, and we feel them in our bodily preparations for such movements. That is, they affect our bodily attunement. This is an aspect of Husserl's phenomenology that Merleau-Ponty develops, and I'll return to this point later.

## 28.3 CONSCIOUSNESS HAS A TEMPORAL STRUCTURE

Not only does Husserl have much to say about intentional, spatial, and aesthetic aspects of experience; he also has a dev0eloped analysis of the temporal nature of consciousness. Experience is structured in such a way that it is not simply in the present, but includes an anticipation of what is about to happen, and a retention of what has just happened. Consciousness, on Husserl's account, has a three-fold structure of retention, primal impression, and protention. The retentional aspect at any moment of consciousness provides a sense of what I have just experienced; primal impression is directed at the present moment; and protention is oriented towards what may happen in the next moment.

Retention and protention involve a double intentional structure (see Figure 28.2). Retention, for example, is primarily directed at the previous phase of consciousness — the just past phase of consciousness that I have just lived through. This just past experience contains in itself a retentional element directed at the just previous phase of consciousness which again includes a further retention, etc. etc. Thus my current awareness includes a sense of the flowing or fading past of my experience, built up in a train of retentions, and this provides me with a sense of self-connectedness. I don't spring into existence anew in every moment, but I experience myself as extended over time. If, for example, I am listening to a melody, I have an immediate and lived sense or feeling (which is not yet a recollective memory) that I have been engaged in this listening process over the course of the last several seconds or minutes. In addition, retention keeps open a second dimension. Since each past phase of consciousness that is retained also involves a primal impression of the object that is experienced I retain a sense of the continuity (or of something changing) in the object that I experience. If I am listening to a melody, then not only do I have a sense that I have been listening to it for some time, I also have a sense of how that melody has developed. Indeed, without this kind of retentional temporal structure, which is built into the perceptual act, we would not be able to perceive melodies at all. We would perceive only the current moment of one note sounding, and then we would have to add something like a recollection of having heard the previous note. The noesis would have to be a complex combination of momentary perceptions supplemented by a succession of memories rather than an actual perception of a developing melody.

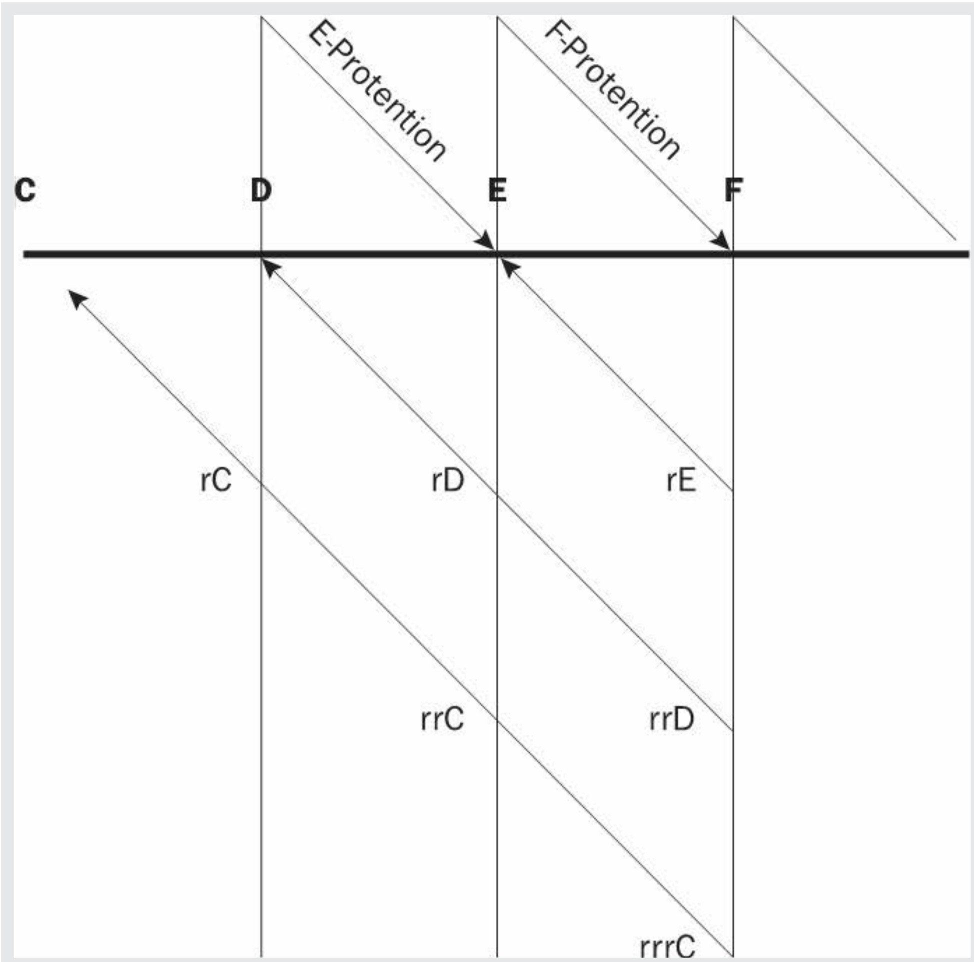



FIGURE 28.2: A diagram of the temporal structure of consciousness. The horizontal line designates a series of sounded notes (C, D, E, F). The vertical lines indicates phases of consciousness, consisting of protentions (above the horizontal), primal impression (where horizontal and vertical meet), and retentions (below the horizontal). The diagonal line (e.g. rC, rrC, rrrC) illustrates how a specific note (C) remains in consciousness in an ever-growing retentional train. (From Gallagher and Zahavi 2012).

Although Husserl doesn't apply this analysis to protention, it is clear that protention has a similar double intentionality. In protention I anticipate both (1) that I will continue to experience something, i.e., that it will be my continuing experience, and (2) that in such experience the object, e.g., the melody, will continue to appear in a similar or changing way. The protentional structure of consciousness provides an account of how I can be either unsurprised — if what I anticipate actually happens as anticipated — or surprised — if what I anticipate doesn't happen in the way I expect it to. In listening to a melody my anticipation may thus be fulfilled as the familiar melody continues as expected, or disappointed if the musician suddenly hits a wrong note, or if the melody is suddenly interrupted by some other event.

This temporal structure of experience is what psychologists call "ecological," which means that it delivers in an ongoing way both a coherent experience of the world, and simultaneously a self-experience of my own coherence as a subject. With respect to objects in the environment, the coherence delivered by the temporal structure of experience is important if, for example, the sense of an object, including the aesthetic effect of the object, depends on experiencing it over time. This pertains not only to temporally changing objects like melodies, but to temporally stable objects like physical things that are either stationary or moving in space. In many cases, for instance, I may have to manipulate or walk around the object, or wait for the object to move in a certain way to gain the full effect of its appearance. For any such object to appear in a coherent way over time, my experience of it must involve the retentional-protentional structure.

## 28.4 FROM DASEIN TO DESIGN

FIGURE 28.3: Martin Heidegger.

*Copyright: pd (Public Domain (information that is common property and contains no original authorship)).*

One of Heidegger's most important insights is that our primary stance towards the world is more like a pragmatic engagement with it than like a detached observation of it. We don't simply open our eyes and look at the objects around us; rather, we are more inclined to grab things and use them. Even if we are just looking, we look at things in terms of their relevance for our pragmatic use. Our intentionality is shaped by this orientation to action. From this perspective the things around us appear as "ready-to-hand" (Zuhanden). Heidegger drives this home with the example of a hammer. For the carpenter who is proficient in her use of the hammer, the hammer is not something that she contemplates or theorizes about; it is rather something that she picks up and uses, and in doing so, the hammer is such that it becomes experientially transparent — that is, it is barely noticeable as it is assumed into a complex involvement in the project of the carpenter. The hammer as such has the status of an instrument, tool, or piece of equipment that serves to support my action or project. More generally, I approach objects, I am in relation to the world, by way of my projects. Heidegger suggests that the hammer takes on a different status if for some reason it breaks, or if it is improperly designed and difficult to use. Suddenly the hammer is an object to be thought about, a problem to be fixed — it becomes something "present-at-hand" (Vorhanden) which gets in the way of my project, or prevents me from moving forward towards my goal.

According to Heidegger, philosophers and scientists have generally, and mistakenly, developed their explanations of the world by taking things as primarily present-at-hand — that is, by treating things as objects to be observed from a detached, objective perspective. Indeed, this sort of objective thinking gets applied not only to things in the environment, but to human agents as if their primary way of existing were to be present-at-hand things; and to Being itself, as if Being itself were an entity, or thing, or a particular being. The human being (or human existence, Dasein) is precisely not a thing. It has a different kind of existence altogether and is primarily an agent, action-oriented, and that, at least in part, defines its existence.

One can get a sense of how this ontological difference, not only between Being and particular entities, but also between objects or things and human agents, has been misconstrued in the history of philosophy and science by looking at Aristotle's analysis of human existence. In the Physics, Aristotle is concerned to provide the proper ontological categories for understanding physical things. In this respect he offers his famous hylomorphic theory. That is, he suggests that every physical entity is composed of matter (hyle) and form (morphe). The latter in the first instance means shape. If we take a particular material and give it a certain shape, we also give it a certain function and make it what it is. The design of an automobile, for example, must meet certain requirements, both materially and morphologically, in order to function as an automobile. The form, the function, and the thing's "whatness" (eidos, essence) define it as such. Aristotle explicates his analysis in terms of the four causes: the material and formal causes, as just defined, and the efficient and final causes. With regard to an automobile, for example, an explanation of the material that goes into it, and its functional shape, plus an account of how these things came together (e.g., a manufacturing process) which is the efficient cause, plus an account of why anyone would want to build an automobile, or what its purpose is, which is the final cause, would provide its complete explanation.

All of this is well and good for explaining physical things and doing physics, at least of the sort that Aristotle envisioned. Aristotle, however, thought this kind of explanation was also useful for explaining other things. Thus, in his Metaphysics, he used precisely the same categories, specifically form and matter, to explain being qua being; and in his Psychology he used precisely these same categories, derived from his analysis of physical things, to define the human agent. This, from Heidegger's perspective, was a mistake. To use explanatory categories from physics to explain Being itself is to reduce Being to a thing-like entity and to ignore the ontological difference between Being and particular entities. And to use explanatory categories from physics to explain human existence is to reduce the human being to an object and, at the very least, to ignore its peculiar kind of agency or its particular way of being-in-the-world.

Dasein, the human being, has the following kind of existence, which is different from any other kind of being. The human being is characterized as always finding itself already in existence, discovering itself already in specific circumstances and moods (an existential characteristic which Heidegger calls Befindlichkeit, disposition, or thrownness). Another way of saying this is that human agents always finds themselves situated. Furthermore, humans are always projecting meaning — they always finds themselves in a world that they are constantly trying to interpret (a characteristic which Heidegger calls Verstehen or understanding).

Human existence is also always intersubjective — humans are always influenced by their encounters with others, in such a way that part of their existential nature is to be-with others (Mitsein or being-with). In these various ways human existence is differentiated from the kind of being that things have. Setting aside the question of whether non-human animals can have this kind of existence, Heidegger clearly thinks that non-human physical things of the sort that we encounter in the environment do not have dispositions, are not attempting to understand, and are not interacting with others in an intersubjective way. The meaning of things, and specifically whether they have a mode of being that is ready-to-hand, like the hammer-in-use, or present-at-hand, like the broken hammer, depend on the fact that humans give them meaning within the context of various projects, most of which arise through intersubjective and social processes.

## 28.5 FROM HAMMERS TO WINE GLASSES

FIGURE 28.4: Jean-Paul Sartre in 1950.



Jean-Paul Sartre, who was influenced by both Husserl and Heidegger, once remarked that with phenomenology one can philosophize about a wine glass. That is, one can begin to think about the purpose of a wine glass within the context of human projects. Why the wine glass is designed in the specific way that it is designed may have something to do with function or style, but it is clear that it is designed for humans and by humans. Like Heidegger, Sartre suggested that we encounter not only the specifics of our own existence in our products and how we use them, but we encounter others in such products. We encounter them implicitly as producers and as organizers of the means of production, but we also encounter them in the utility provided by such products. We encounter others in the practices that employ such objects. In this respect, viewing an object from both a present-at-hand perspective and a ready-to-hand perspective is essential for design. That is, a designer must treat her product as something Vorhanden — as something that is shaped in a certain material that facilitates function — but also as something Zuhanden — something best suited to the use of human agents. This leads us directly to considerations about the human body since it is normally with a human hand that we reach out to grasp a wine glass, and it is only with a human mouth that we take the sip.

## 28.6 EMBODIED AND ENACTIVE DESIGN

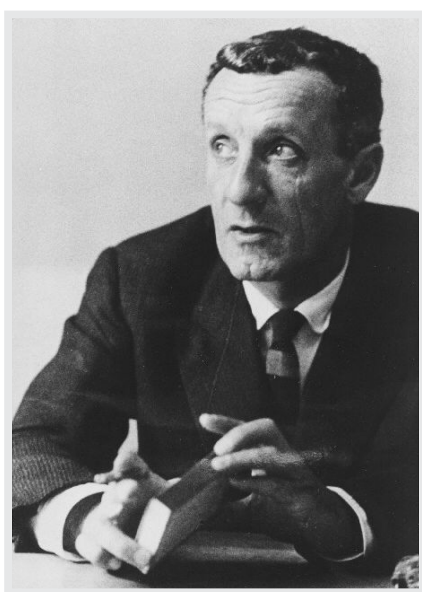


Figure 28.5: Maurice Merleau-Ponty.

In the work of Merleau-Ponty we find greater emphasis on the body than in either Husserl or Heidegger. In contrast to the traditional denigration of the body found in western philosophy from Plato to Descartes, and in certain forms of gnostic religions, where mind, or soul, or spirit dominates, Merleau-Ponty treats the body as the perceiver, the knower, the agent. Intentionality is primarily a motor intentionality. We make sense of the world through our bodily actions. Indeed, what remains implicit in Heidegger becomes explicit for Merleau-Ponty, namely, that our primary relation to the world, as pragmatic, and as ready-to-hand, is primarily an embodied relation where our hands and motor systems necessarily play an essential role.

Merleau-Ponty takes over the basic phenomenological distinction between lived body (Leib) and objective body (Körper) from Husserl. Recently this distinction has been phrased as the body-as-subject (or we might say as-agent) vs the body-as-object. We engage with the world, and we encounter others intersubjectivly, for the most part as embodied agents rather than as passive observers or detached minds. Our bodies are not primarily objects — either for ourselves or for others — but agentive bodies that express their subjectivity in their postures, movements, gestures, actions, and expressions. My body as perceiver and agent is in-the-world, engaged in projects, and participating with others in the making of meaning.

Merleau-Ponty has inspired recent work on enactive and extended conceptions of cognition. On the enactive view, our perception of the world is essentially linked to our movements and our motor possibilities (Varela et al, 1991). Perception is pragmatic in the way indicated by Heidegger, and it's not just a matter of the brain processing information or representing objects in the environment on a kind of internal map. Our perception is based on specific sensory-motor contingencies (see O'Regan and Noë 2001). I see and understand an object in terms of what I can do with it. In this regard Merleau-Ponty exploits Husserl's insights about the kinaesthetic dimension associated with perception. My perception is informed by my ability (or lack of ability) to move. Husserl and Merleau-Ponty also talk about the “I can” intentionality of perception. The fact that an object appears in the reachable (peripersonal) part of the environment where I can reach it, for example, as opposed to an unreachable (extrapersonal) part of the environment has an influence on its relevance, its valence, and how I perceive it. This phenomenological insight has been confirmed by neuroscientific studies that show different activation patterns under these different conditions (see e.g. Caggiano et al. 2009). Our pragmatic, action-oriented way of being-in-the-world is also reflected in the activation of “canonical” neurons that are activated both when we reach out to grab a tool or object, and simply when we perceive that tool or object (Grezès et al. 2003). We perceive the world in terms of the possibilities for action that it offers.

Accordingly, perception is guided by what J. J. Gibson has called ‘affordances’. For example, I see a chair as affording me the possibility of sitting. That clearly depends on the way it is designed and how that design is related to my body. An object that fails to provide a flat horizontal surface doesn’t afford sitting; nor does a chair that is three inches tall. A well-designed chair is not a sitting-affordance at all for a lion or for an animal or robot that does not have flexible joints. Affordances depend on specific relations or negotiations between objective shape and the shape and flexibility of the agentive body.

## 28.7 EXTENDING THE MIND INTO THINGS

Like the enactive view, the extended mind hypothesis (Clark and Chalmers 1998; Clark 2008) argues that cognition is not simply “in the head”, but depends on extra-neural bodily and environmental elements. Cognition extends to processes in the environment, to processes involved in our use of technical instruments that support our cognitive activities. Our handy use of pencil and paper not only facilitates our calculations, it can, in part, constitute our cognitive process. My memory can supervene on the use of a PDA as much as it can on my hippocampus; my solutions to navigational problems can incorporate the use of GPS; my correct spelling can be the result not only of my biological memory but also of my use of a spell check feature on my laptop. There are debates within the philosophical community about how far we can take such claims about the extended nature of cognition, but most theorists accept that tools and technologies can facilitate our cognitive processes. This also applies to the way that we design our environments. The way that we set out instruments in an operating room can facilitate the performance of the surgeon; the way we design a cockpit can facilitate the decisions of the pilot; the way we furnish a classroom can support learning. All such design details can also work the other way. That is, our cognitive tasks can be thwarted, our imagination stifled, our problem solving slowed, our judgments clouded when our environment is ill designed, just as much as when our tools and technologies start to interfere with our thinking.

One essential aspect of both enactive and extended conceptions of cognition is a specification of how precisely we are coupled with tools, technologies, and environments. That is, the cognitive process is in some cases constituted and in some cases simply facilitated depending on the nature of the body-environment coupling. Once again this will depend both on various bodily habits, the condition of the body, and the nature of the bodily practices engaged, and on the arrangement and condition of the environment. These aspects are measurable and explainable in terms of dynamic systems theory (see, e.g., Chiel and Beer 1997). They can also be explored experimentally using methods of neurophenomenology (Varela 1996).

Let me mention that much of what I have just indicated about perception and cognition can be repeated in regard to intersubjective processes and social arrangements. For phenomenology following Merleau-Ponty my ability to understand another person involves an embodied set of processes that are sensory-motor and contextual rather than primarily inferential or simulational. Specifically, I am not primarily a third-person observer of others; rather, I engage and interact with them in socially defined contexts. My understanding of others is also facilitated by the kind of environments that we find ourselves in, the kind of interactions that such environments (including artifacts, tools and technologies) allow. Environments can be built in ways that can either promote or prevent communication, and in this sense physical environments are also social environments and have direct relevance for intersubjective interactions.

## 28.8 THE DESIGN OF INTELLIGENT SYSTEMS

In the 1970s Hubert Dreyfus relied on phenomenological insights from Heidegger and Merleau-Ponty to develop a critique of strong AI — that is, the kind of artificial intelligence projects that aimed to construct machines with human-like intelligence. Dreyfus presented a general critique of computational models of cognition. Thinking of the brain as a computer that simply processes information in discrete operations not only misses a number of important aspects of human cognition; it entirely misconstrues the nature of cognition. For example, it misses the Heideggerian and enactive conception that cognition is for action and that the human is pragmatically involved in the world. Important for this view is the fact that the world offers highly complex and ambiguous situations that require a kind of interpretation that is not possible if one has only a set of rules to follow. If AI can create machines that function in extremely circumscribed contexts where a machine or robot can do what they do by following a set of formal rules, this kind of intelligence generally does not transfer to other contexts, and especially changing contexts where one needs to recognize without rules what kind of situation one is confronted with. These kinds of problems typically are related to the frame problem in AI. Dreyfus's position continues to be debated in recent books and articles (see Wheeler 2005; McDowell 2007).

Dreyfus also appealed to Merleau-Ponty's work on embodiment, and this has special reference to robotics. If in fact cognition is embodied from the bottom up, then the design of robots with a set of top-down executive controls will have limited success in performing pragmatic tasks in ambiguous situations or in interacting with human agents. This idea motivated Rodney Brooks (1991) to re-think AI and robotics to address basic sensory-motor problems, and to attempt to build intelligence from the bottom up. For Dreyfus, to design a robot with human-like intelligence would require the robotic body to be human-like, not just in appearance, or in its ability to pragmatically engage in the world, but to have bodies more or less like ours, with the possibility of social interaction and acculturation.

## 28.9 YOU AND I, ROBOT

Researchers in advanced robotics are attempting to design and build autonomous social robots that will be able to seamlessly and reliably interact with humans in specific situations. A huge amount of research funding is currently being invested in human-robotic interaction (HRI). There has been some progress in building robots that can learn to respond to (and with) certain emotional expressions, and to respond to (and with) specific gestures. There are still serious limitations in designing robots that can operate in the kind of ambiguous situations that typify human projects and interactions (see Gallagher 2007; Gallagher 2012b). This motivates both philosophical and practical questions about what precisely is required in a robot if it is to be able to engage in something close to the kind of interaction that characterizes human-human relations, even if only on a pragmatic level.

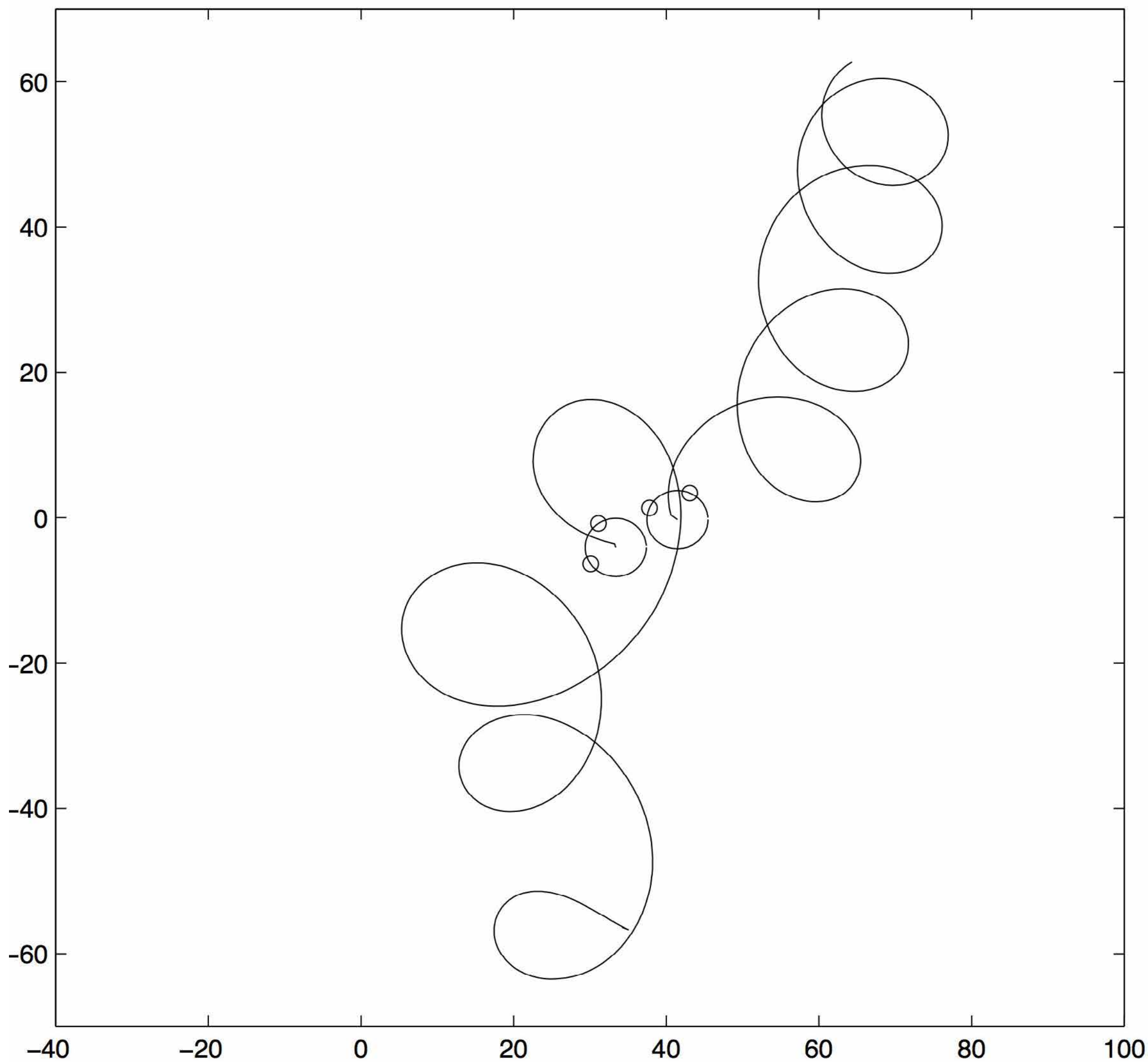
Dreyfus is right to suggest that we need to have a correct understanding of human cognition if we want to design intelligent systems. We also need a good understanding of human-human interaction if we want to design a robot capable of interacting with humans in smooth and problem-free ways. This means that we need the right theory of social cognition, and this is another area in which phenomenology has been challenging received doctrines and offering its own constructive account.

In contrast to the standard "theory-of-mind" (ToM) accounts of social cognition, which focus on "mindreading" construed as observation-based inference or the running of simulation routines, phenomenological approaches to social cognition emphasize the embodied and enactive aspects of our interactions with others. Our understanding of others is based on sensory-motor, perceptual and interactional abilities in rich and meaningful (and, in most cases, socially defined) contexts that are also defined by artifacts, tools, and technologies. In most everyday situations, I can get a sense of what you want to do in the way you are doing it. That is, I don't have to attempt to infer or simulate your mental states. In many cases your intention is obvious to me because I am not simply observing your behavior in an off-line (third-person) mode; rather I am engaged with you in an on-line (second-person) interaction. Moreover, as research in developmental science and phenomenology suggests, in our interactions I understand the meanings of your actions (and gestures and expressions, etc.), enactively, that is, in terms of social affordances, in terms of my possible responses to your actions. In many cases, also, the particular situation (the physical setting, the social environment) does some of the work. The meaning of a certain gesture or a certain action is specified by the social or physical situation in which it is enacted; the meaning of the same set of movements may be specified differently in a different situation.

In regard to designing autonomous robots that are capable of genuine social interaction, standard ToM theories suggest that we need only create specific mechanisms inside the individual robot that will provide it with the ability to infer, simulate, or mindread. In contrast, phenomenologically inspired interaction theory claims that in many circumstances interaction itself (the dynamic, enactive engagement of two or more individuals which is not reducible to the actions of the individuals qua individuals) constitutes social cognition (de Jaegher et al 2010). Just as in dancing the tango, the emergent dynamic phenomenon, the tango itself, is something more than an addition of individual contributions. In social contexts, as in dance contexts, and as predicted by the phenomenological analysis of the temporality of experience, the timing specifics of interactions are essential. This suggests that even if a workable social robot were designed on standard ToM principles to support inferential or simulative mindreading, for it to be capable of smooth and reliable interaction with humans it would already have to be a robot that behaves sufficiently like us so as to engage in embodied interaction.

If this is true, then at a minimum we need to design robots that have similar sensory-motor abilities and that can recognize the significance of the same kind of rich and meaningful contexts in which humans act, rather than machines capable of more precise inferential calculations or abilities to run more powerful simulations. In effect, social understanding has to be already implicit in the robot's actions and in its abilities to see others in terms of how they can interact with them (i.e., as social affordances).

This approach places extraordinary demands on robotic design. Is it even possible to build such a robot? Not only should we not rule anything out apriori, ongoing research in evolutionary robotics provides minimal instances of such interactive machines. Di Paolo (2000), for example, evolved agents that were able to achieve coordination through interaction. Two robots, whose only task was to locate each other and remain close as they moved through a large space, used simple auditory signals and rotating motor behavior, to set up a specific sound pattern that differentiated between self and non-self, and simplified what would otherwise be a complex recognition problem.


60
40
20
0
-20
-40
-60
–40
–20
0
20
40
60
80
100

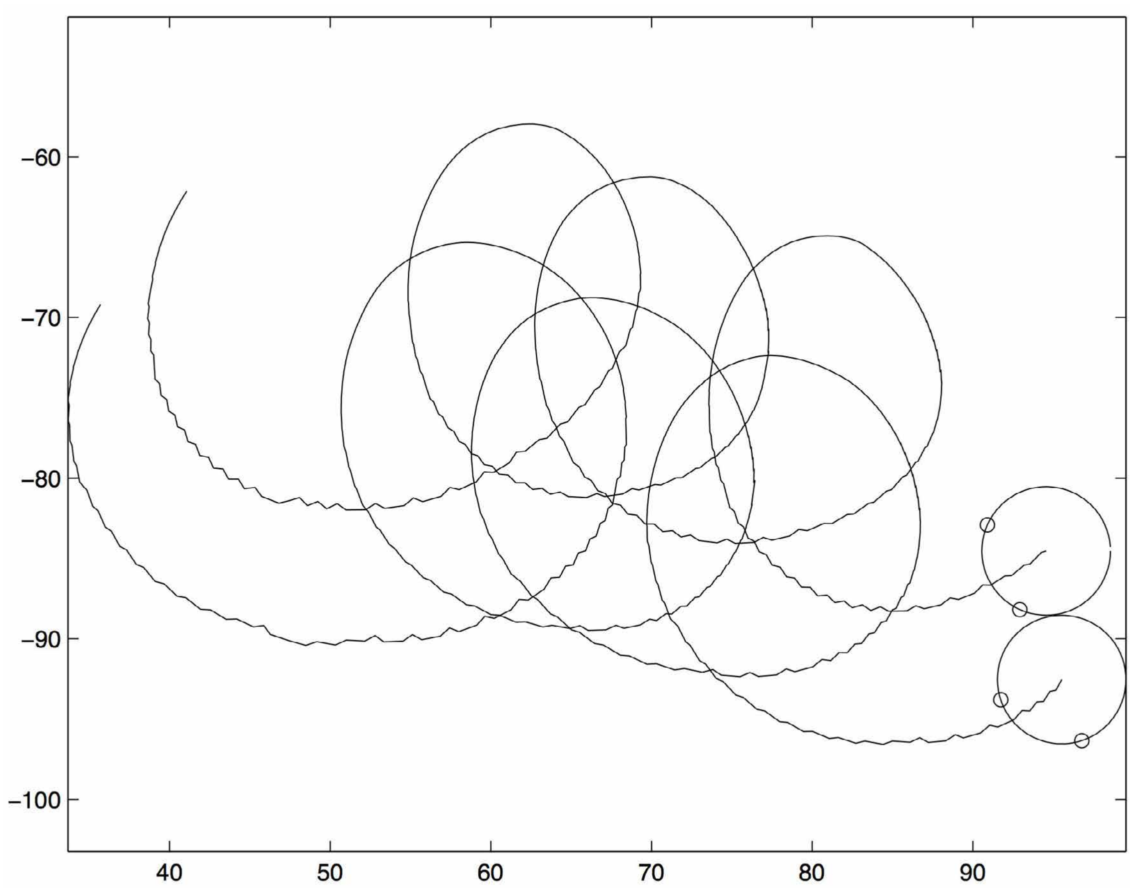





FIGURE 28.6 A-B: Left: Trajectories of approaching agents. Right: Motion of agents during period of coordination. From Di Paolo 2000.

Importantly, when an individual agent was presented with a recording of its partner from the previous successful interaction, that is, when it was presented with precisely the same data that it had previously processed in an interactive way, it was unable to reproduce its own behavior due to the fact that the other robot's recorded actions and "responses" were non-contingent and mis-timed. One-sided coordination is not achievable, which suggests the important contribution of interaction itself. Similar results have been found in other experiments involving artificial agents (Auvray et al 2009; Di Paolo et al 2008). Such experiments in evolutionary robotics can supplement phenomenological methodology in order to identify the enactive principles involved in minimal social interaction (see Froese and Gallagher 2010; Froese and Gallagher 2012). The challenge is to see if these principles scale up to the kind of non-minimal behaviors that will characterize full-out human-robotic interactions in worldly environments.

## 28.10 CONCLUSION

Although phenomenology originates as a philosophical enterprise, it has practical application in a number of fields, including design and human-robotic interaction. As a study of our intentional, temporal, and lived experience it is directly relevant to design issues. It suggests that the intentional structure of consciousness as informed by bodily movement and kinaesthetic sensation will shape our phenomenal and aesthetic experiences. It points to our action-oriented way of being-in-the-world as determining how we experience the things around us, and it emphasizes the important role that embodiment plays in perception and cognition. Phenomenology explores the ways that our physical and social environments, including the things and instruments in such environments matter for experience, cognition, problem solving, and for shaping our intersubjective and social interactions.

## 28.11 ACKNOWLEDGEMENTS

My work on this chapter was s0upported by a Research Fellowship at the Kolleg-Forschergruppe Bildakt und Verkörperung at Humboldt University, Berlin (2012). Also thanks to the ongoing support of the Marie Curie Initial Training Network, Towards an Embodied Science of Intersubjectivity (TESIS). Marie Curie Actions, European Commission Research for support of my research on intersubjectivity, and the Humboldt Foundation fellowship: The Anneliese Maier Research Award [Anneliese Maier-Forschungspreis] at the Ruhr-Universität Bochum (2012-17).

## 28.12 REFERENCES

Auvray, Malika, Lenay, Charles and Stewart, John (2009): Perceptual interactions in a minimalist virtual environment. In New Ideas in Psychology, 27 (1) pp. 32-47

Brooks, Rodney A. (1991): Intelligence without representation. In Artificial Intelligence, 47 (1811) pp. 139-159

Caggiano, Vittorio, Fogassi, Leonardo, Rizzolatti, Giacomo, Thier, Peter and Casile, Antonino (2009): Mirror neurons differentially encode the peripersonal and extrapersonal space of monkeys. In Science, 324 (5925) pp. 403-406

Chiel, Hillel J. and Beer, Randall D. (1997): The brain has a body: adaptive behavior emerges from interactions of nervous system, body and environment. In Trends in Neurosciences, 20 (12) pp. 553-557

Clark, Andy (2008): Supersizing the Mind: Embodiment, Action, and Cognitive Extension (Philosophy of Mind). Oxford University Press, USA

Clark, Andy and Chalmers, David J. (1998): The Extended Mind. In Analysis, 58 (1) pp. 7-19

Dreyfus, Hubert L. (2007): Why Heideggerian AI failed and how fixing it would require making it more Heideggerian. In Artificial Intelligence, 171 (18) pp. 1137-1160

Dreyfus, Hubert L. (2002): "Intelligence without representation" Merleau-Ponty's critique of mental representation The relevance of phenomenology to scientific explanation. In Phenomenology and the Cognitive Sciences, 1 (1962) pp. 367-383

Dreyfus, Hubert L. (1992): What Computers Still Can't Do: A Critique of Artificial Reason. The MIT Press

Froese, Tom and Gallagher, Shaun (2010): Phenomenology and Artificial Life: Toward a Technological Supplementation of Phenomenological Methodology. In Husserl Studies, 26 (2) pp. 83-106

Gallagher, Shaun (2007): Social cognition and social robots. In Social Cognition, 3 (3) pp. 435-453

Gallagher, Shaun (2006): How the Body Shapes the Mind. Oxford University Press, USA

Gallagher, Shaun and Zahavi, Dan (2012): The Phenomenological Mind: An Introduction to Philosophy of Mind and Cognitive Science, 2nd edition. Routledge

Gibson, James J. (1979): The Ecological Approach to Visual Perception. New Jersey, USA, *Lawrence Erlbaum Associates*

Heidegger, Martin (ed.) (1982a): *The Basic Problems of Phenomenology, Trans. Albert Hofstadter. Indiana University Press*

Heidegger, Martin (1989): *Ideas Pertaining to a Pure Phenomenology and to a Phenomenological Philosophy: Studies in Phenomenology of the Constitution. Transl. R. Rojcewicz and A. Schuwer. Springer*

Heidegger, Martin (ed.) (1982b): *Ideas Pertaining to a Pure Phenomenology and to a Phenomenological Philosophy: First Book: General Introduction to a Pure Phenomenology. Transl. F. Kersten. Springer*

Heidegger, Martin (1996): *Being and Time (translated by Joan Stambaugh and Dennis J. Schmidt). State University of New York Press*

Jaegher, Hanne De, Paolo, Ezequiel A. Di and Gallagher, Shaun (2010): *Can social interaction constitute social cognition?. In* Trends in Cognitive Sciences, *14 (10) pp. 441-447*

Mcdowell, John (2007): *What Myth?. In* Inquiry, *50 (4) pp. 338-351*

Merleau-Ponty, Maurice (1962): *Phenomenology of Perception: An Introduction. Routledge*

O'Regan, J. Kevin and Noë, Alva (2001): *A sensorimotor account of vision and visual consciousness. In* Behavioral and Brain Sciences, *24 (5)*

Paolo, Ezequiel A. Di (2000): *Behavioural coordination structural congruence and entrainment in a simulation of acoustically coupled agents. In* Adaptive Behavior, *8 (1) pp. 25-46*

Paolo, Ezequiel A. Di, Rohde, Marieke and Iizuka, Hiroyuke (2008): *Sensitivity to social contingency or stability of interaction? Modelling the dynamics of perceptual crossing. In* New Ideas in Psychology, *26 (2) pp. 278-294*

Varela, Francisco J. (1996): *Neurophenomenology: A methodological remedy for the hard problem. In* Journal of Consciousness Studies, *3 (4) pp. 330-349*

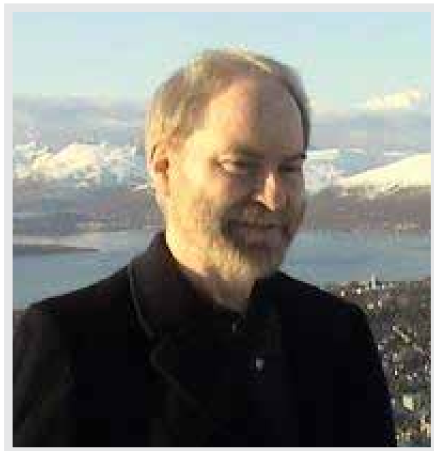
Varela, Francisco J., Thompson, Evan and Rosch, Eleanor (1991): *The Embodied Mind: Cognitive Science and Human Experience. Cambridge, Massachusetts, MIT Press*

Wheeler, Michael (2007): *Reconstructing the Cognitive World: The Next Step. A Bradford Book*

Wheeler, Michael (2005): *Reconstructing the Cognitive World: The Next Step (Bradford Books). The MIT Press*

## About the author

### Shaun Gallagher



© *Shaun Gallagher*

Personal Homepage:
[http://pegasus.cc.ucf.edu/~gallaghr](http://pegasus.cc.ucf.edu/~gallaghr)

Current place of employment:
Department of Philosophy, University of Central Florida

Shaun Gallagher is the Lillian and Morrie Moss Professor of Excellence in Philosophy at the University of Memphis. He has a secondary appointment at the University of Hertfordshire (UK) and is Honorary Professor of Philosophy at the University of Copenhagen (Denmark). He's held visiting positions at the Cognition and Brain Science MRC Unit at the University of Cambridge, the Ecole Normale Supériure in Lyon, and the Centre de Recherche en Epistémologie Appliquée (CREA), Paris, and he is currently Visiting Fellow at the Kolleg-Forschergruppe Bildakt und Verkörperung at Humboldt University, Berlin. He holds the Humboldt Foundation Anneliese Maier Research Fellowship (2012-17) and is PI on grants to conduct research on intersubjectivity and institutions (Marie Curie Foundation) and the aesthetic and spiritual experiences of astronauts during space travel (Templeton Foundation). His publications include How the Body Shapes the Mind (Oxford, 2005); The Phenomenological Mind (with Dan Zahavi, Routledge, 2008), and as editor, the Oxford Handbook of the Self (Oxford, 2011). He's editor-in-chief of the journal Phenomenology and the Cognitive Sciences.

## Your notes and thoughts on chapter 28

Record your notes and thoughts on this chapter. If you want to share these thoughts with others online, go to the bottom of the page at:
http://www.interaction-design.org/encyclopedia/phenomenology.html

**NOTES:**

# Chapter 29

# Formal Methods

*by Alan J. Dix.*

The use of Formal Methods in human-computer interaction dates back to its earliest days as a growing discipline, including Phyllis Reisner's use of BNF to specify user interfaces in 1981 (Reisner 1981) and the author's own first paper on the topic at the first British HCI Conference in 1985 (Dix and Runciman 1985).

To some extent, Formal Methods sit uneasily within interaction design. Human beings are rich, complex, nuanced, engaged in subtle and skilful social and material interactions; reducing this to any sort of formal description seems at best simplistic. And yet that is precisely what we have to do once we create any sort of digital system: whether an iPhone or an elevator, Angry Birds or Facebook, software is embedded in our lives. However much we design devices and products to meet users' needs or enrich their experiences of life, still the software inside is driven by the soulless, precise, and largely deterministic logic of code. If you work with computers, you necessarily work with formalism.

Formal Methods sit in this difficult nexus between logic and life, precision and passion, both highlighting the contradictions inherent in interaction design and offering tools and techniques to help understand and resolve them.

In fact, anyone engaged in interaction design is likely to have used some kind of formal representation, most commonly some sort of arrow and sketch diagram showing screens/pages in an application and the movements between them. While there are many more complex formal notations and methods, these simple networks of screens and links demonstrate the essence of a formal representation. Always, some things are reduced or ignored (the precise contents of screens), whilst others are captured more faithfully (the pattern of links between them). This enables us to focus on certain aspects and understand or analyse those aspects using the representation itself (for example notice that there are some very long interaction paths to quite critical screens).

### What is Formal?

*As with all words, "formal" is used to mean different things by different people and in different disciplines. In day-to-day life, formal may mean wearing a dinner jacket and bow tie or using proper language. That is, formal is about the outward form of things — a formal greeting may belie many emotions beneath the surface.*

*Taken strongly, formalism in mathematics and computing is about being able to represent things in such a way that the representation can be analyzed and manipulated without regard to the meaning. This is because the representation is chosen to encapsulate faithfully the significant features of the meaning.*

*from* **(Dix 2003)**

## 29.1 KINDS OF FORMAL METHODS IN HCI

There are very many kinds of formal or semi-formal notations, models and techniques used within HCI. One way to categorise them is by what gets represented:

**users** — Various forms of cognitive models have been used to analyse interaction, from *cognitive architectures* such as SOAR (Laird 2008) or ACT-R (Anderson 2005) to motor level models. The latter have continued to be surprisingly influential, however they are largely limited to numerical fitting of data with a few exceptions such as Eslambolchilar's work using *control theory* to model the cybernetic interaction between human movements and digital devices (Eslambolchilar and Murray-Smith 2010); (Eslambolchilar 2006). *Task analysis* and *task models* (Diaper and Stanton 2003) are also aimed at analysing or recording the behaviour of users, whether on existing systems or planned systems. Some task modelling systems take into account multiple users (e.g. CTT (Paterno 1999)), and there has been some modelling within the CSCW community (e.g. (Ellis 1994); (Dix et al. 2000) ); however, modelling at the group or social level is rare.

**system** — The other side of the interaction is the device or system that the user is interacting with. Many techniques have been applied to modelling the system behaviour including formal grammars, state machines, specification languages from theoretical computer science, and plain sums! While the focus of this kind of model or representation is the thing (computer, consumer device), this does not mean the user has been ignored. Rather, the needs of the user are expressed in the properties examined, or in the choice of what aspects to model.

**world** — Representations of the context of interaction can be very insightful: users are interacting with devices in a physical context, and their digital interactions may well have physical consequences, whether it is the controls of an aeroplane, or simply printing. However, models that take this into account are surprisingly rare. One example has been the use of space syntax, techniques developed within architectural theory in order to understand people's movements and visitation patterns in urban or office spaces (see Section 29.4.1), and the author's own work has included modelling of interactive art installations (see Section 29.4.2). *Context-aware interfaces* (see Chapter 14) also often build some form of internal model. Later in this chapter, we will look at modelling of physical devices that include aspects of both system and world.

The representations of the user, while certainly formal, are typically not what is thought of as 'Formal Methods' and are likely to be dealt with in other chapters. Also, as noted above, representations of the world are rare, so the rest of this chapter focuses principally on Formal Methods applied to the system.

However, this itself has many variants:

**at what architectural level do we represent the system?** — This may be focused on presentation details, for example to assess the visual discrimination of items, or Fitts' Law timings of user actions; it may be at the dialogue level expressing the order of potential user actions and user responses, or at a deeper level analysing or modelling the underlying functionality insofar as it impacts on the user.

**at what level of abstraction do we represent a system?** — We may have a very concrete model of a particular system design, or we may opt for a more generic/abstract model in order to investigate some property irrespective of the particular system in which it is found. An example of the latter is analysis of undo functionality, discussed in Section 29.3.3.

**for what purpose are we representing the system?** — We may create a formal representation to be part of the *execution* of a running or prototype system. Alternatively, we may use the formal representation to perform automated or hand *analysis* (e.g. average number of keystrokes between different system states). Finally, it is often the case that it is the process of formalising that gives the analyst a deeper *understanding* of the system being studied.

## 29.2 JUST SUMS - AD HOC CALCULATIONS

Straightforward mathematical calculations are everywhere in HCI. These range in complexity. At the simplest level is simple arithmetic, for example, in the GOMS keystroke-level model (Card et al. 1980); (Card et al. 1983). The models behind *information foraging theory* are more complex, using differential equations (Pirolli 2007). In areas such as visualisation, information retrieval, and graphics, mathematics is again central.

Even small 'back of the envelope' calculations can be surprisingly effective in helping one to understand a problem. We will look at two such examples: screen typing and menu depth, then at the more complex mathematics behind "five users are enough".

### 29.2.1 On screen typing

Some years ago, comparisons of graphical vs. more textual interfaces (and they do still exist!) often used the idea that graphical displays have high 'bandwidth'. This obviously should be interpreted in terms of visual perception, not just raw pixels per second, but for output it seems fairly reasonable. But what about input, do screen buttons and icons increase the input bandwidth? In fact, a quick Fitts' law calculation shows that no matter what the number and size of the screen buttons, a reasonable typing speed is always faster (see box). The difference is that whereas the lexicon of the keyboard is fixed and has to be interpreted by the user in any context, GUIs can be contextual, showing appropriate actions (if you know any information theory, this is a form of adaptive compression). Notice that a small mathematical argument can lead to a design perspective.

Revisiting these calculations, many years later, they are perhaps more relevant in the face of on-screen keyboards in smart-phones and pad devices. To some extent, the early calculations are borne out in recent studies of iPad on-screen keyboards, which have found ratios more like 2:1 for keyboard to screen typing.

**Back of the Envelope — keyboard vs. screen typing**

Compare keyboard with screen for rate of entry measured in bits per second.

Keyboard:

Take typing times from KLM times quoted in (Dix et al. 2004).

nos targets – 64 keys

good typist – 9 keys per sec.

rate = $9 * \log_2(64) = 54$ bps

Screen:

Screen width W with items size S on it. The average distance to target is half the width. To make calculations easier, assume a square screen and that the screen is completely filled with targets.

Fitts' Law – $0.1 \log_2 ( D/S + 0.5 )$

$D = W/2$

nos items – $(W/S)^2$

$$\text{rate} = \frac{\log_2 ( (W/S)^2 )}{0.1 \log_2 ( W/2S + 0.5 )}$$

$$\approx \frac{2 \log_2 ( W/S )}{0.1 \log_2 ( W/S )}$$

$= 20$ bps

So, screen clicking is nearly three times slower than typing!

### 29.2.2 Optimal menu depth

Often the 7+/−2 rule (Miller 1956), which is about working memory, is mistakenly over applied. One example is for menu systems, and you may well have seen suggestions that the number of menu items per screen (e.g. on a web page) shouldn't exceed 7+/−2. On a touch screen, large targets are easier to hit, again suggesting that small numbers of larger targets are a good idea. However, the fewer menu items on a single screen, the more menu levels are required to find particular content. Let's assume there are N items in total and you choose to have M menu items per screen. The depth of the menu hierarchy, d, is given by:

$$d = \log N / \log M$$

If we look at a single display, the total time taken will be the time taken to physically display the screen and the time taken for the user to select the item, all times the number of levels:

$$T_{total} = ( T_{display} + T_{select} ) \times d$$

Using Fitts' law for $T_{select}$ and the formula for d, we get:

$$T_{total} = ( T_{display} + A + B \log M ) \times \log N / \log M$$

$$= ( (T_{display} + A) / \log M + B ) \times \log N$$

Notice that the effect of the increased number of screens exactly balances the gains of larger targets and that the only factor that varies with the number of menu items is the per screen costs (Tdisplay+ A). This suggests that the more items per screen the better. Look at virtually any portal site, and you'll see that practical experience has come to the same conclusion!

In fact there are extra factors to consider; for very small targets, Fitts' law starts to break down, which puts lower limits on target size. Also errors are very significant as this causes wasted screen displays, so smaller numbers of well-explained items may be better. For larger numbers of items, a further factor sets in, the time taken for the user to locate an item on the display. There is evidence that for linear menus, the actual select time is closer to linear than logarithmic. Redoing the calculations shows that this visual search becomes the limiting factor for larger screens, leading to a maximum menu size depending on refresh time (which is still much larger than 7+/-2 for most cases!) However, good design of screen organisation and sub-headings can push this visual search back towards the logarithmic case (see (Larson and Czerwinski 1998)). For WAP with small scrolling displays, the figures are again different.

Notice that being precise forces us to make assumptions explicit, and also, focusing on the critical factors helps us look for potential design solutions.

### 29.2.3 Five users are enough

Most people working in HCI and interaction design have heard "five users are enough", whether as a justification of a small study, or to refute one with fewer than five. In fact, this is grossly over-used and misapplied, probably more so than Miller's 7+/-2. Here we'll just look at the underlying mathematical model, but see (Dix 2011) for a detailed analysis of why "five" is not a sufficient answer to "how many users are enough?"

In fact, the roots of "five users are enough" originally lay in an analysis of empirical data by Nielsen and Landauer (Nielsen and Landauer 1993). They gathered data from user testing in a number of medium to large software projects. The data they gathered included the number of users studied in each cycle of development, the number of problems found by each user, and whether these were shared with other users. Also, and crucially, they measured the costs of performing each user study and the costs of creating a fresh iteration of the software.

They then used mathematics developed in software engineering to model program debugging, treating user interface problems discovered by user studies, or heuristic evaluations as analogous to coding bugs uncovered during software testing. Each bug (user interface defect) is regarded as equally likely to be found during any period of testing and totally independent of other bugs. The overlap between bugs found during different test runs can be used to estimate the likelihood of finding a single bug, and then this can be used, *inter alia*, to estimate the likely number of as yet undiscovered bugs. This is similar to mathematics used by biologists when doing capture/recapture studies to assess survival rates in populations of wild animals.

The analysis (given the assumptions) shows that the number of new problems found decreases exponentially with the number of additional users, and this can be used to create a cost-benefit graph where the maximum (averaged over different kinds of projects) turned out to be around 5 for heuristic evaluation ... although closer to 3 for user testing (largely because they were deemed more costly). After this, it was more cost-effective to perform an iterative development step (which is assumed, counter to previous assumptions, to 'reset the clock' for user interface problem discovery).

As noted, this result has been widely misapplied (Dix 2011), but we can draw two general lessons:

- mathematical models can be incredibly influential
- it is important to understand critical aspects of these models when applying the results, notably the underlying assumptions
- the formality of the models makes it easier to explore the precise nature of these underlying assumptions, and therefore possible to assess the scope of applicability

## 29.3 DETAILED SYSTEM SPECIFICATION

### 29.3.1 Dialogue notations - what to do when

In day-to-day life, dialogue is a conversation between two or more parties, usually, but not necessarily, cooperative.

In user interfaces, *dialogue* refers to the structure of the interaction, the syntactic level of human–computer ‘conversation’. Dialogue forms the middle of three layers of the user interface as identified in the early days of user interface development, especially in the Seeheim model (Pfaff and Hagen 1995):

**lexical** (*presentation* in Seeheim) – the shape of icons, actual keys pressed

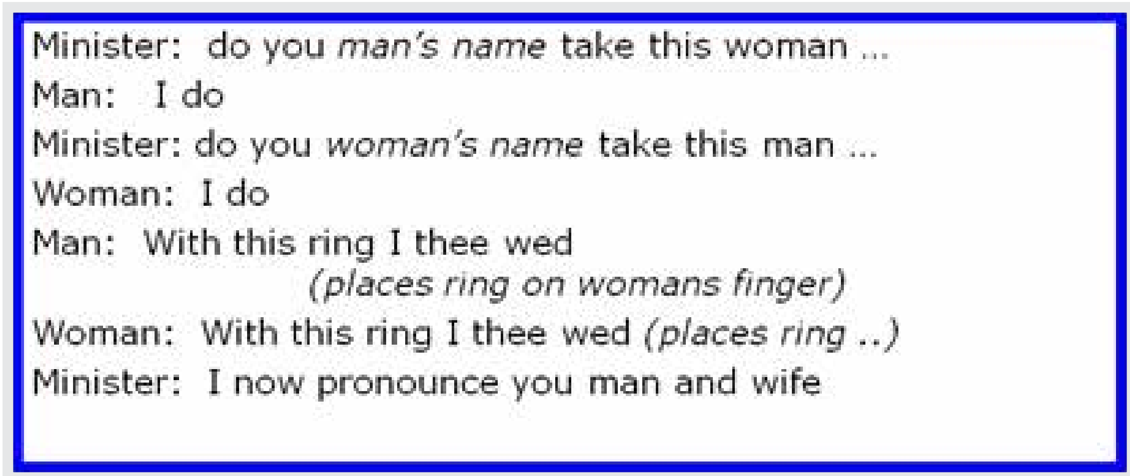
**syntactic** (*dialogue*) – the order of inputs and outputs

**semantic** (*functionality*) – the effect on internal application/data

Of course, compared with human dialogue, human-computer dialogue is very constrained. However, some human-human dialogues are formal too. Think of a wedding service; it is a sort of script for three parties, which exhibits many of the properties of user interface dialogue:

- It specifies order.
- Some are contributions fixed – “I do”.
- Others are variable – “do you man’s name …”.
- The instructions for the ring are concurrent with saying the words “with this ring …”.

The author often asks people to go through the words when giving tutorials about dialogue. Imagine you have just done this, spoken the words to a complete stranger. Does it mean you are married to her/him? Of course not, the words are empty without meaning; they only carry legal weight if said in the right place, with a marriage licence, etc. They are purely the syntax of marriage, with none of the semantics.


Minister: do you *man's name* take this woman ...
Man: I do
Minister: do you *woman's name* take this man ...
Woman: I do
Man: With this ring I thee wed
*(places ring on womans finger)*
Woman: With this ring I thee wed *(places ring ..)*
Minister: I now pronounce you man and wife



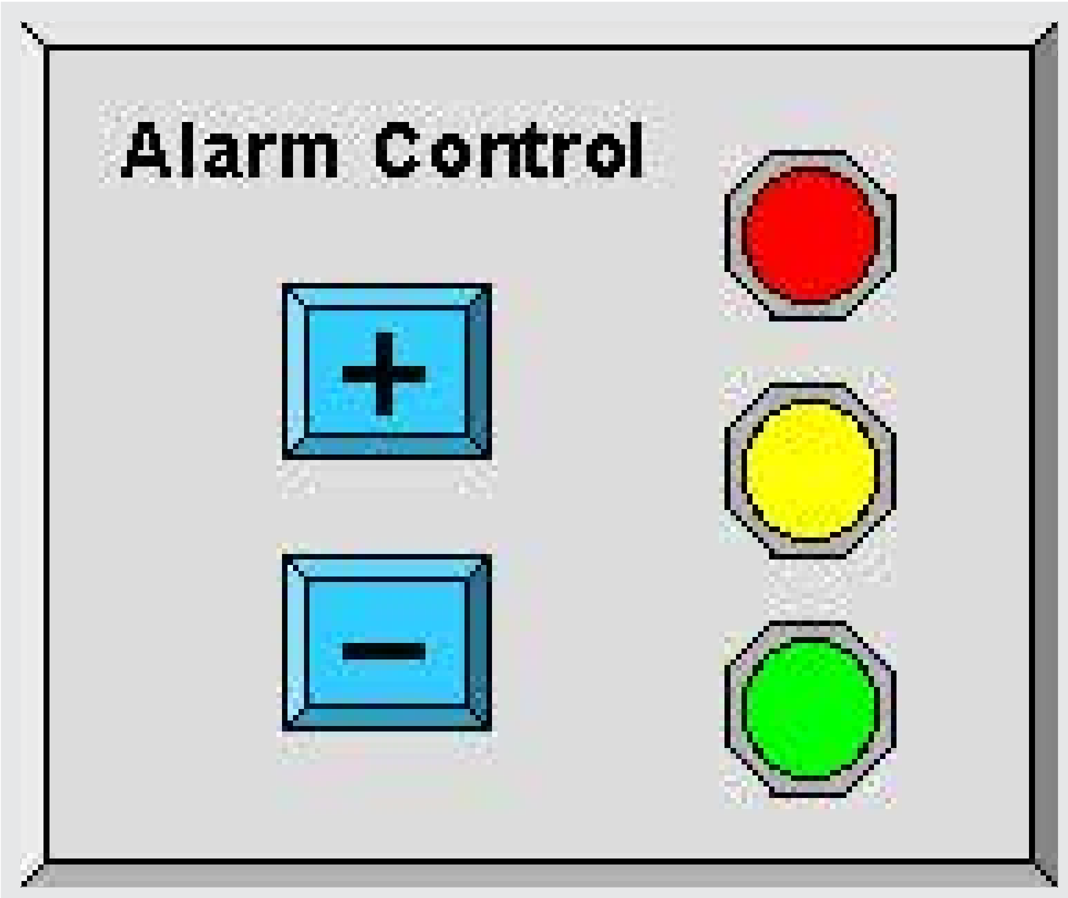
Imagine you are in a wedding service, but when asked, the woman says "I don't"? Real dialogues often have alternatives. The wedding script doesn't cater for this; it assumes that people will work things out for themselves when unexpected things happen. However, when describing human–computer dialogues, we cannot assume the computer will be able to cope with the unexpected, so we have to consider all other alternatives, even if some of them might be 'do nothing'.

Sometimes, formal human dialogues do more clearly deal with these alternatives. In a trial, when the judge asks whether the defendant pleads guilty or not guilty, the process of the trial depends on the defendant's response.

Because people cope, the focus in formalised human dialogues is on normative responses, the standard trial process doesn't cope with the Queen walking in and saying "off with her head". Similarly, while specifications of human–computer dialogues need to be more complete, they will often ignore certain unlikely things, such as the user standing on the keyboard!

Let's see what you can do with dialogue notations and analysis.

Imagine you are analysing the security control panel at a top secret establishment (Figure 29.1.A). It has two buttons labelled '+' and '–' that control the current alarm state. The alarm state can be at three levels, denoted by three lights: green, amber or red, and the buttons increase or reduce the level.



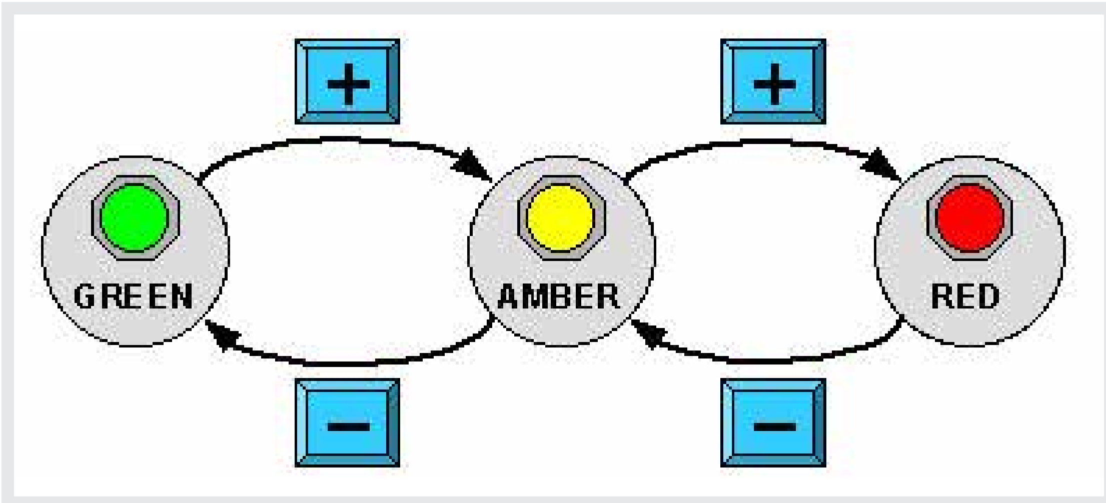





Figure 29.1 A-B: Top secret control panel: (A) control panel (B) state transition network.

The image in Figure 29.1.A is the presentation level, what the panel looks like. A common way to specify the dialogue is using some form of *state transition network* (STN). Each major state or mode of the system is drawn as a circle, or sometimes as a thumbnail of the critical part of the visual display. The effects of the possible user actions (in this case the '+' and '–' buttons) are shown by the way they transform the states. Figure 29.1.B shows a state transition network for the control panel.

Just looking at the STN – without knowing what it is about – we can ask some questions: what does the '+' button do when the state is red, what does the '–' button do when the state is green? This is a formal analysis – we don't need to know what the STN means, but we can see just from the pattern that something is missing. In many applications, such as a program control for a television, we would probably want the '+' button to cycle back round from red to green, but in this application – an operator under stress perhaps – we would probably not want to risk accidentally changing from red alert to green. The formal analysis tells us *that* something extra is needed, the contextual understanding tells us *what* it is.

In this example, we have examined the diagram by hand, but an advantage of formal representations is that they can often also be analysed automatically. There are various forms of this.

*Model checkers* have been used in HCI research to verify whether certain desired usability properties are true (e.g. (Campos and Harrison 2001)). Note that these need to use quite sophisticated ways to cut down the total number of alternatives, as the number of possible paths through even a simple interface grows rapidly: if there were only 10 possible actions (icons, menu selections, valid keystrokes), then there are 100 possibilities for two actions, 1000 for three, … Also the number of possible states grows rapidly. Imagine we have one option, say whether a font is bold, italic, bold-italic or neither: that has four states. Now, imagine a second option, say five possible font sizes at 9, 10, 11, 12 or 14 point: the number of possible font states is not 4+5, but 4x5=20. If we then add underline or not, we have 40 states. Because of this, model checkers use a combination of exhaustive and symbolic evaluation, but are still limited in the complexity of the systems they can analyse.

In fact, there are many safety-critical cases where the number of options is too large for human analysis, but can be dealt with by methods such as model checking, or other mathematical techniques such as graph analysis or matrix algebra. The FIT Lab at Swansea applies these techniques to medical instruments, allowing them to work out the likelihoods of different kinds of errors (Cairns et al. 2010).

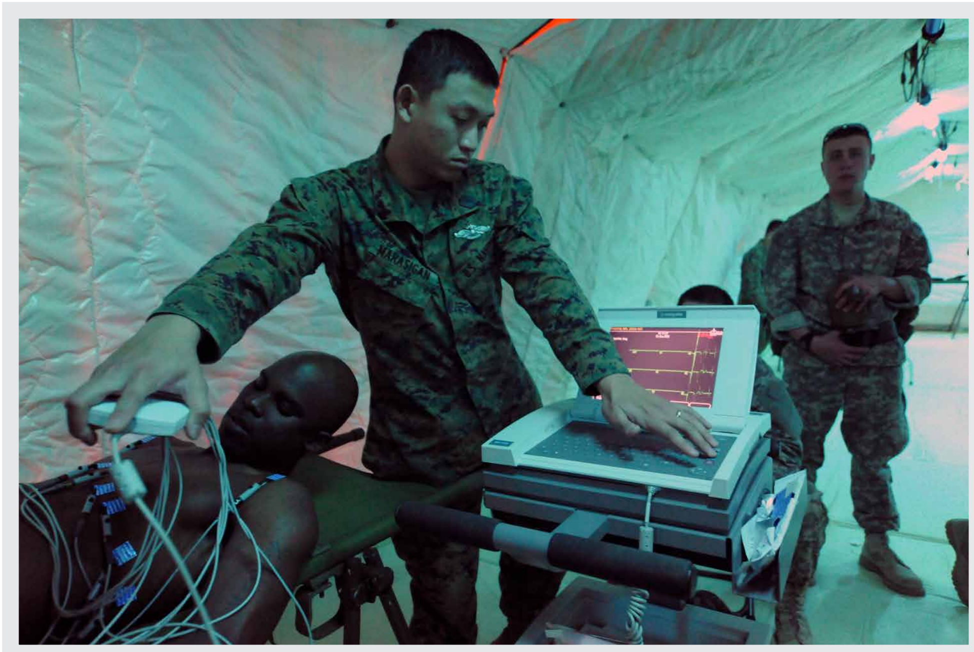


Figure 29.2: Medical instruments is one example where the number of options may be too large for human analysis, but can be dealt with by methods such as model checking.

*Courtesy of Bruce Cummins, US Navy. Copyright: pd (Public Domain (information that is common property and contains no original authorship)).*

### 29.3.2 Executable system models

One of the uses of early dialogue notations was as part of prototype or deployed user interfaces. When coding user interfaces, whether point-and-click GUIs or web-based systems, it is easy to get lost in the effects of single actions, "when this button is pressed, what happens next". Various notations were common in the early days of *user interface management systems* (UIMS), including formal languages such as *BNF*, rule-based transition systems, and *state transition networks* (STNs). The last of these have been most common probably due to their graphical nature (even if usually translated into textual form for execution), and have continued to be used, for example, the Arrowlets framework for

JavaScript-based user interfaces (Phang et al. 2008). However, other formalisms continue to be used, for example, Petri Nets have been used extensively, especially in work surrounding safety critical design such as air-traffic control (Navarre et al. 2001).

It is rare to see anyone outside a research lab use any of the most precise dialogue specification techniques, but link diagrams, such as those produced by Denim (Lin et al. 2000) and similar tools, are common, including some that are executable at least as prototypes, although these are not normally regarded as 'formal'. Also teams using UML are likely to use State Charts, although applied at a low level rather than for the dialogue.

- *Model-based user interfaces* start modelling at a higher level than dialogue taking user task models and then refining these in stages to eventually obtain running systems. The CAMELEON reference model (Calvary et al. 2002) captures this at four levels:
- *Final User Interface* (FUI) – The actual interface running on a particular platform and language with platform-specific widgets.
- *Concrete User Interface* (CUI)– Here a Java Swing JButton and an HTML button (<input type="submit"/>) would be both described as a generic "Graphical 2D push button".
- *Abstract User Interface* (AUI) – The interface is now abstracted in a modality independent way as 'abstract interaction objects' (AIO); so that a "control AIO" may be a 2D button on screen, or a physical button on the device. Also relationships between AIOs are defined in terms of spatial and temporal constraints, but not in terms of precise layout.
- *Domain/Task Concepts* – At this level, the interface itself is all but forgotten. Instead, the focus is on what is wanted; for example

"load a file" which then at a more concrete level may be rendered as a button launching a file selection dialogue.

This has been used in a number of systems including UsiXML (Limbourg et al. 2005); (www.usixml.org accessed March 2011), an XML-based collection of user-interface notations and tools.

### 29.3.3 Abstract system models

Both dialogue models and executable models are about a specific system being analysed or developed. However, Formal Methods can be applied to the analysis of broad issues that affect classes of systems.

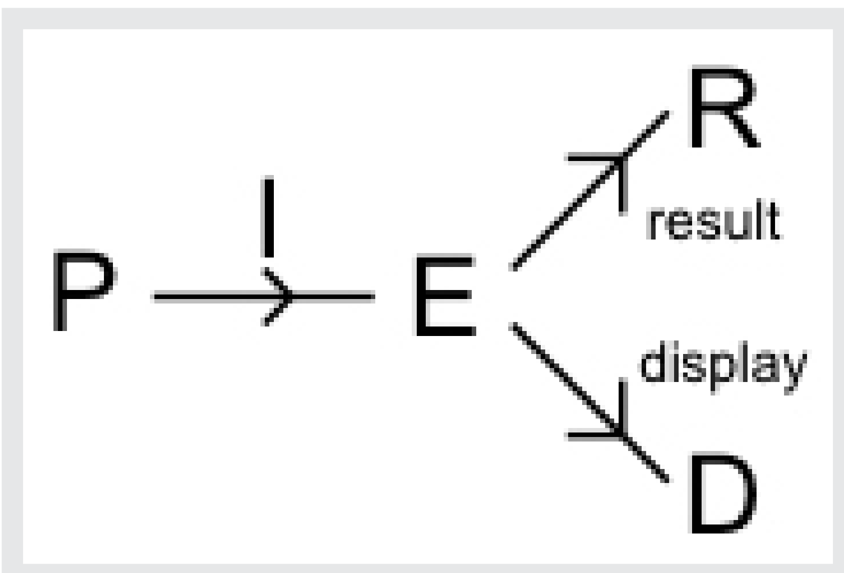



FIGURE 29.3: The PIE model.



One of the earliest abstract models of this kind was the PIE model, developed by myself and colleagues at York in the mid-1980s (Dix and Runciman 1985). The PIE model treats an interactive system in input–output terms. The user input is some sequences of basic commands (e.g. individual mouse clicks, keystrokes). For historical reasons, the set of sequences of commands was denoted P (standing for 'program'). The traces of user commands are interpreted by the system (I denotes this function) giving rise to 'effects' (E) of various kinds from the internal system state to the bell sounding. Often there were two classes of effects: the more permanent results (R), such as the printed document and the moment-to-moment display (D). The latter whilst called 'display' also included effects such as haptic or aural feedback.

Variations were developed to allow analysis of issues of time, multi-window displays, and non-deterministic interactions (Dix 1987); (Dix 1990). Also, the basic PIE model was sometimes recast in terms of system states and an update function 'doit' as this was sometimes easier used to discuss and analyse problems. One of the nice things about using a formal model is that instead of arguing about which representation was the best, it was possible to write down the way they mapped on to one another:

$$\text{doit}(\, c, I(p)\,) = I(\, p\, c\,)$$

and then we were able to use whichever was best for any job.

While very minimal, the PIE model allowed us to give precise definitions of a number of properties connected with the observability of the system, driven in no small part by the popularity of the idea of WYSIWYG ('what you see is what you get') at the time.

One of the simplest is:

$$\exists\ \text{observe}: D\text{->}\ R\ \text{st.}\ \forall\ e \in\ E : f(\, \text{display}(e)\,) = \text{result}(e)$$

This says that there exists (∃) a function (called here 'observe') from displays to results, such that for all (∀) states of the interactive system (effects), the function applied to the current display gives you the current result. Or in other words, you can tell what you are going to get by looking at the current screen.

Of course, there are often parts of a document that are off screen, so more complex variants of these were specified to be able to say, “there is always some set of navigation commands that let you see enough to be able to tell the result”.

These kinds of properties were then used by those with notations and tools for particular systems as things to check, either by hand, or using model checkers or other automated analysis tools.

Undo systems have been extensively studied, partly because they were a particularly hot topic at the time this work was first being done, partly because they have nice algebraic properties that are easy(ish) to handle using formalisms, and partly because they have a nasty edge case which make them hard to deal with or error prone to deal with informally.

One of the earliest results (Dix 1991) was to show that it is impossible to have a ‘universal undo’, that is a single ‘undo’ button, which would undo the previous action no matter what it was, including undo itself. At the time, there were various systems that had variants of toggle undo (swopping between two states), and some of these (it was claimed) obeyed the universal undo property.

The universal undo property can be framed (using the ‘doit’ representation):

$$\forall\ s \in E,\ c \in Cu\ \ doit(\ undo,\ doit(\ c,\ s\ )\ ) = s$$

Let’s go through the impossibility proof:

**1)** consider any state, s, of the system and any two commands a and b. Imagine the effect of either doing a or doing b. These will yield two states (typically different) $s_a$ and $s_b$.

$$s_a = doit(\ a,\ s\ ),\ s_b = doit(\ b,\ s\ )$$

**2)** consider applying undo in states $s_a$ and $s_b$, first state $s_a$

$$doit(\ undo,\ s_a\ ) = doit(\ undo,\ doit(\ a,\ s\ )\ )\ \{\ expand\ \}$$

= s { universal undo }

by the same argument doit( undo, $s_b$ ) = s also

that is undo in each case gets us back to the original state, just as we would expect

**3)** finally consider applying undo in state s. As well as being the original state, it was also the result of applying undo in state sa (step 2). That is, we might have done the sequence a, undo, undo. Mathematically:

doit( undo, s ) = doit( undo, doit( a, $s_a$ ) ) { from (2) }

= $s_a$ { universal undo }

however, equally the same argument applies if we think of b:

doit( undo, s )= doit( undo, doit( a, $s_b$ ) ) { from (2) }

= $s_b$ { universal undo }

So, the two states must be the same: $s_a = s_b$

As the argument started in step (1) by choosing any state and any pair of commands, this means wherever we are in the system, the effect of any command is always to do the same thing! It is as if there were only one button in the system, every keystroke, mouse click, or thumb on the back with a hammer does exactly the same thing. Furthermore, because undo of undo gets us back to where we started, it turns out there are at most two states in the system ... the light switch is the most complicated system that can have the universal undo property.

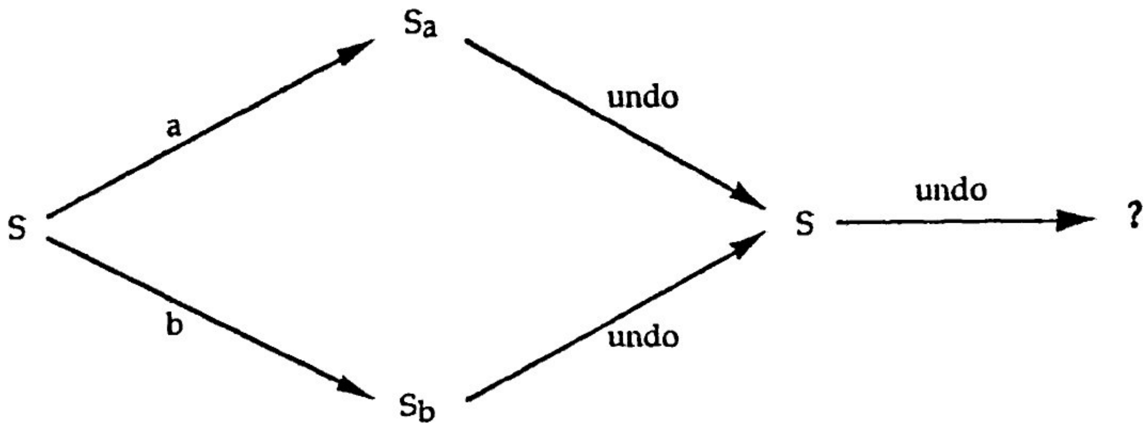



FIGURE 29.4: Space syntax colouring.



In fact, real undo systems never obey this property (they are more complicated than light switches!), but instead they are variants of two kinds of undo:

*toggle undo* where undo toggles you back and forth between the previous and current state, but you can never go back more than one step

*stack undo* where the system remembers a stack of previous states. Normal commands add to the stack (a partial history), and undo/redo navigates up and down the stack (like back and forward in a web browser). However, as soon as you do any normal command in the middle of doing undo/redo, the stack is chopped off where you are and the new state added instead.

Formal analysis of these alternatives led to analysis of the 'danger' points for stack undo (when a slip during deep undo/redo cycle can lose massive amounts of work), and it also provided a proof that toggle and stack undo were the *only* mechanisms that satisfied basic properties for undo (Dix et al. 19997). This analysis used a quite complex branch of mathematics called Category Theory, which is especially powerful for 'arrow chasing' proofs around diagrams like those in Figures 29.3 and 29.4.

Furthermore, the structural similarities between undo and browser back meant that the same analysis techniques were able to be applied to this also, distinguishing the fact that 'back' whilst apparently simple (like undo) was in fact subtly, but importantly, different in different kinds of systems.

There were many extant single-user undo systems at this point, but multi-user editing was still new. Formal analysis of multi-user undo allowed the potential problems and solution strategies to be identified before the first such systems were created. When group undo systems were created, they encountered exactly the problems and solutions that the formal analysis had uncovered.

This exemplifies one of the core strengths of Formal Methods, the ability to be able to analyse problems before actual systems exist.

## 29.4 MODELS OF THE WORLD

As noted in Section 29.1, models of the world are more rare than models of computer systems, but can be very valuable. There are no clear categories of these, often being ad hoc to meet particular needs, so this section just gives three examples of quite different kinds of formal models.

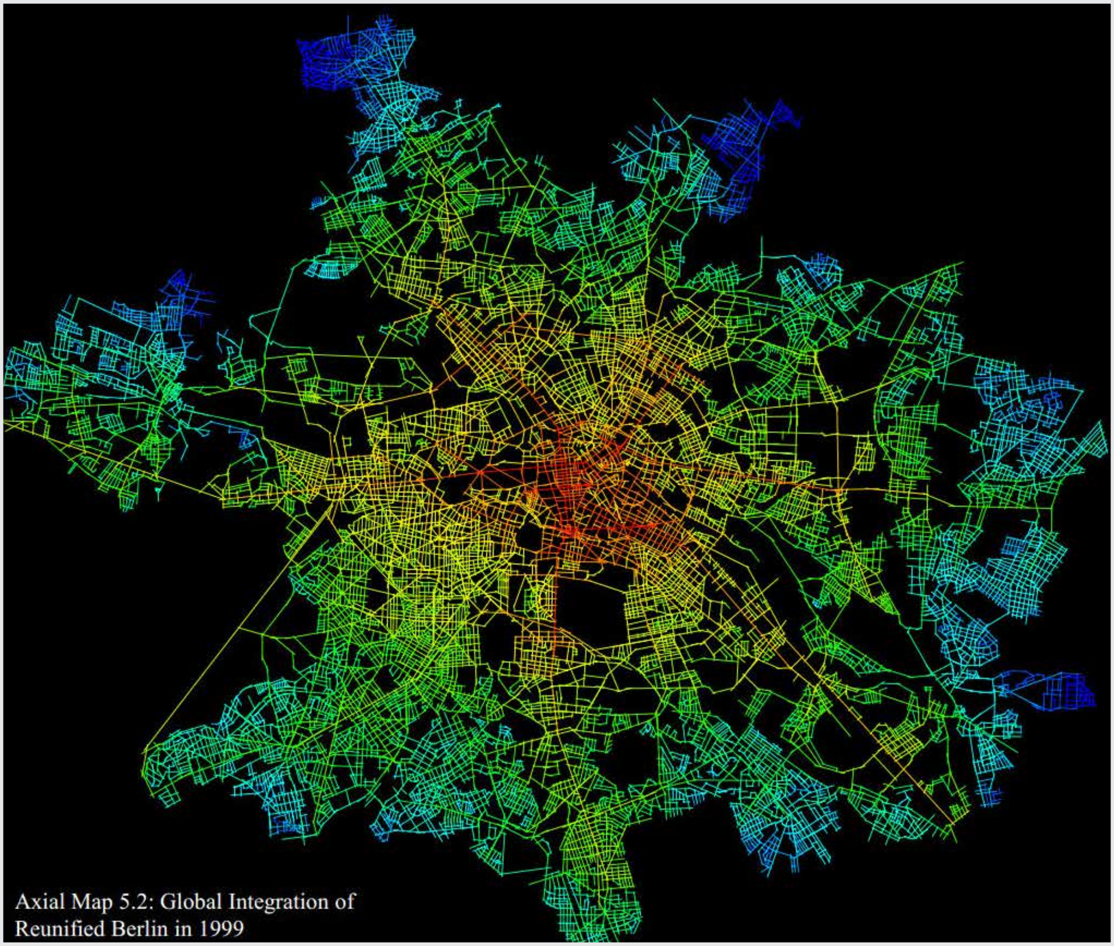
### 29.4.1 Space syntax - modelling movement and significance in the world

Space syntax was initially developed by Hillier at the Bartlett School of Architecture in London (Hillier 1999). There were various works that attempted to describe issues of coherence or legibility of urban spaces, perhaps most influential Lynch's 'image of the city' (Lynch 1960), but Hillier was looking for a more rigorous, theoretical account and for more rigorous tools.

When looking at urban road networks, the obvious measure of 'distance' is not the crow-flies string stretched between points; unless you are in a tank or a helicopter, you cannot simply go through or over buildings. Instead, one is likely to measure the length of the path along roads (often called the ‘Manhattan block’ distance), maybe taking into account one-way systems for cars. If one were being a little more sophisticated, one might look at time taken. For pedestrians, this would be close to the walking distance, but for cars this might distort the map as city centres tend to be busy.

One of the key insights of space syntax is that none of these metrics captures the human sense of the city, and that central to this is line of sight. If two landmarks are within line of site to one another, they are in a sense linked and become ‘close’ in our mental model of the city. Furthermore, it is the major turnings that form psychological markers along the way. Space syntax uses these turnings as metric of distance. If you have to make 5 turnings to get from point A to B, then their space syntax distance is 5.

With this as a metric, locations in a city can be analysed to obtain a centrality measure: a place is more ‘central’ if its maximum/average distance to all other places is lower. With a crow-flies or Manhattan-block distance, this is likely to be close to the geometric centre of the city, but with the space-syntax turnings metric, it may end up in different places. Figure 29.5 A-B show the Berlin streets - before and after the unification - coloured according to the space syntax approach. The white line on Figure 29.5 B is the Berlin wall. Notice how the wall means that the most central area from a space syntax point of view (coloured red) is not in the centre of the city before the unification.



Axial Map 5.2: Global Integration of Reunified Berlin in 1999

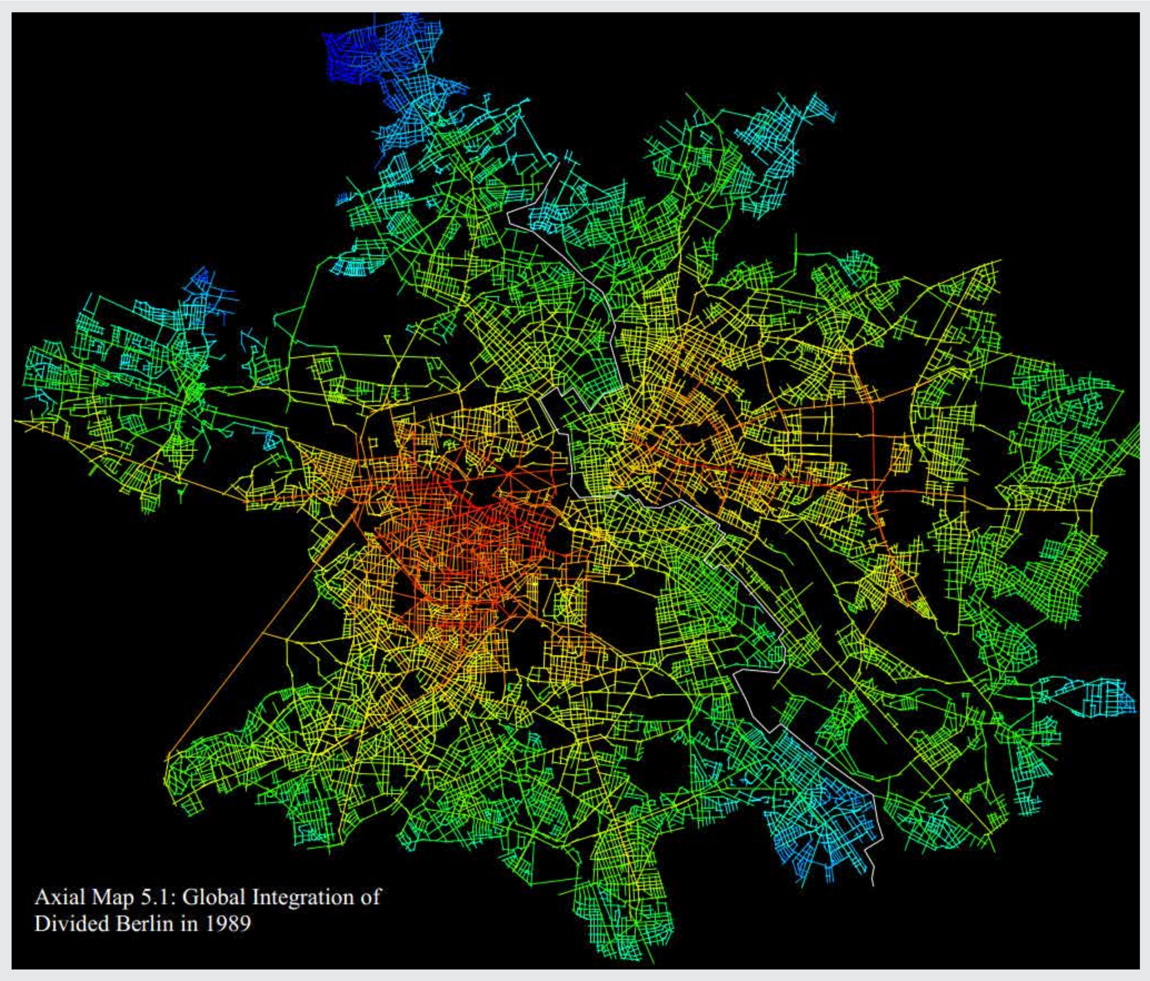





*Figure 29.5 A-B: Space syntax colouring of Berlin - before and after the Berlin wall.*

This far, this is simply drawing pretty pictures, but the remarkable thing is that these measures of centrality can be compared with other measures, for example, the density of shops, restaurants, and other public meeting places. When this is done, the two are found to correlate very closely, that is the space-syntax measure drawn solely from the geometry of the city is a very strong predictor of the real, human heart of the city.

When the Berlin city analysis is repeated with the wall removed, the 'central' areas change accordingly in the space syntax analysis. Within a few years of unification, the busy areas of Berlin shifted; some areas that were previously quiet, residential areas became business or shopping areas. These shifts were perfectly in line with the space syntax predictions.

Similar analysis techniques are used for buildings, and they enable the planner to determine suitable places for meeting rooms, coffee areas, etc.

Space syntax has been used in HCI in two ways.

First, it has been used in order to understand digital interventions in physical spaces. For example, in the Cityware project, it was used to understand mobile application use including computational simulations of movements (Kostakos et al. 2010).

The second use has been in the generation of virtual environments. In the Tower project (Prinz et al. 2004), inhabited virtual spaces were built where users moved amongst representations of documents and folders, rather like those found in cyberpunk novels. Whereas a real city evolves, these are created spaces, and so space-syntax was used generatively to build 'intelligible' information spaces.

### 29.4.2 Belief and time - modelling an artistic performance

Art seems even more distant from Formal Methods than usability, and yet formalism has been used to make sense of artistic performance and art installations.

In certain kinds of performance art, the nature of who is the performer and who is the audience can be deliberately problematic. Imagine, in the midst of a crowded city square a group of performers, all wearing white hats, start to move in unison; this may not be obvious to those around them, but may be visible to someone on a high building. Similarly, if you are in a shopping mall and someone starts to behave oddly, it may take a while before you realise she is a mime artist. In the case of certain TV shows, the reaction of unwitting bystanders becomes part of the spectacle.

One key aspect of dramatic theory is the idea of the 'frame' bounded by time and (maybe virtual) space: whether it is the stage during a theatre performance, or the space around the juggler in the city square. To constitute a performance, the audience needs to be aware that there is a performance frame, and the performers need to know that their audience knows that there is a frame. This is all about belief, and so a form of belief logic has been used to model this.

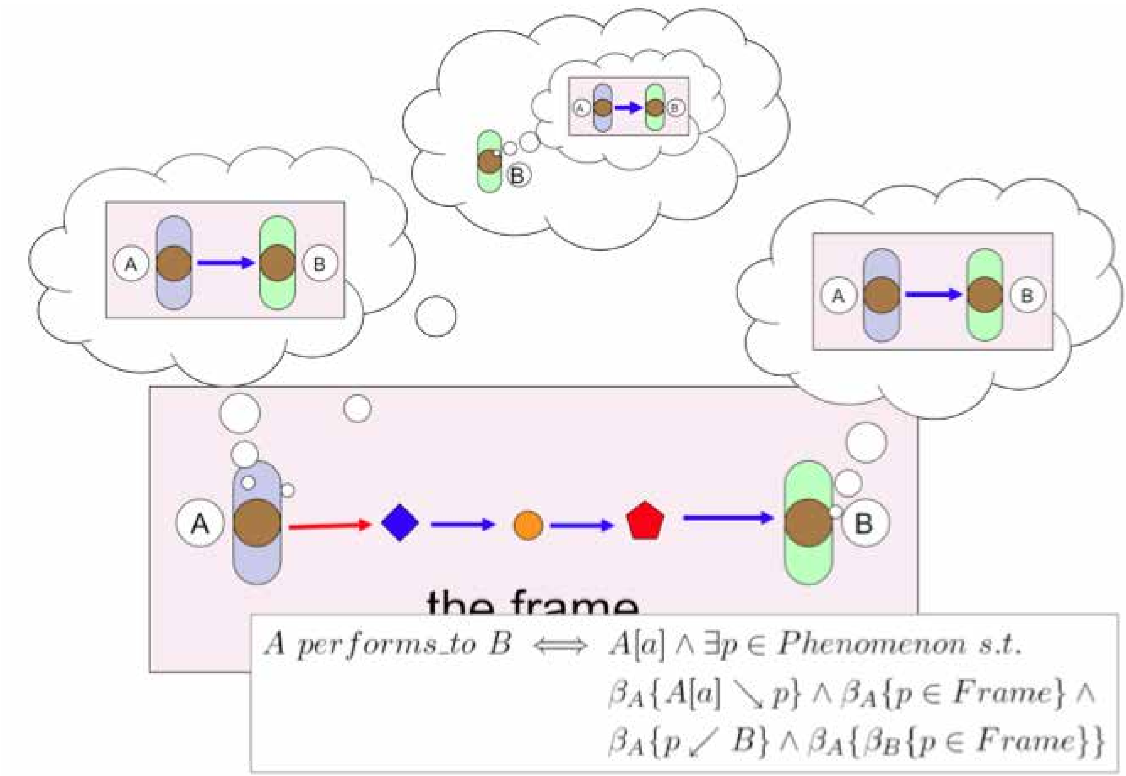



Figure 29.6: This figure shows the situation graphically and in logic. Alison performs to Brian if (i) Alison believes her actions are (potentially indirectly) observable by Brian, (ii) Brian believes the actions are part of a dramatic frame, (iii) Alison believes that Brian believes that the actions are part of the frame. This complex belief chain (A believes that B believes that A is doing something) is hard to think about, and so formalising helps to keep track of who knows what when. This was then used to analyse a particular installation Deus Oculi, which deliberately makes some of the interconnections between parts obscure. The analysis showed that the key 'aha' moment of the installation is precisely when a particular transition of beliefs occurs.

| | actual situation | $B_A$ (A's beliefs) | $B_B$ (B's beliefs) |
|---|---|---|---|
| 1. A and B come to the exhibit | mirror ∈ Frame<br>screen ∈ Frame<br>mirror ↷ screen | mirror ∈ Frame<br>screen ∈ Frame<br>+ knows B's beliefs | mirror ∈ Frame<br>screen ∈ Frame<br>+ knows A's beliefs |
| 2. A looks at mirror, B looks at screen | mirror ↪ A<br>screen ↪ B<br>A ↷ mirror | mirror ↪ A<br>screen ↪ B<br>A ↷ mirror | mirror ↪ A<br>screen ↪ B<br>A ↷ mirror |
| 3. B sees A in screen | A ↷ mirror ↷<br>screen ↪ B | | |
| 4. B infers relationships | | | mirror ↷ screen |
| 5. A looks at mirror, B looks at screen | mirror ↪ B<br>B ↷ mirror<br>mirror ↷ screen<br>screen ↪ B | screen ∈ Frame | B ↷ mirror ↷<br>screen ↪ A<br>+ knows A's beliefs |

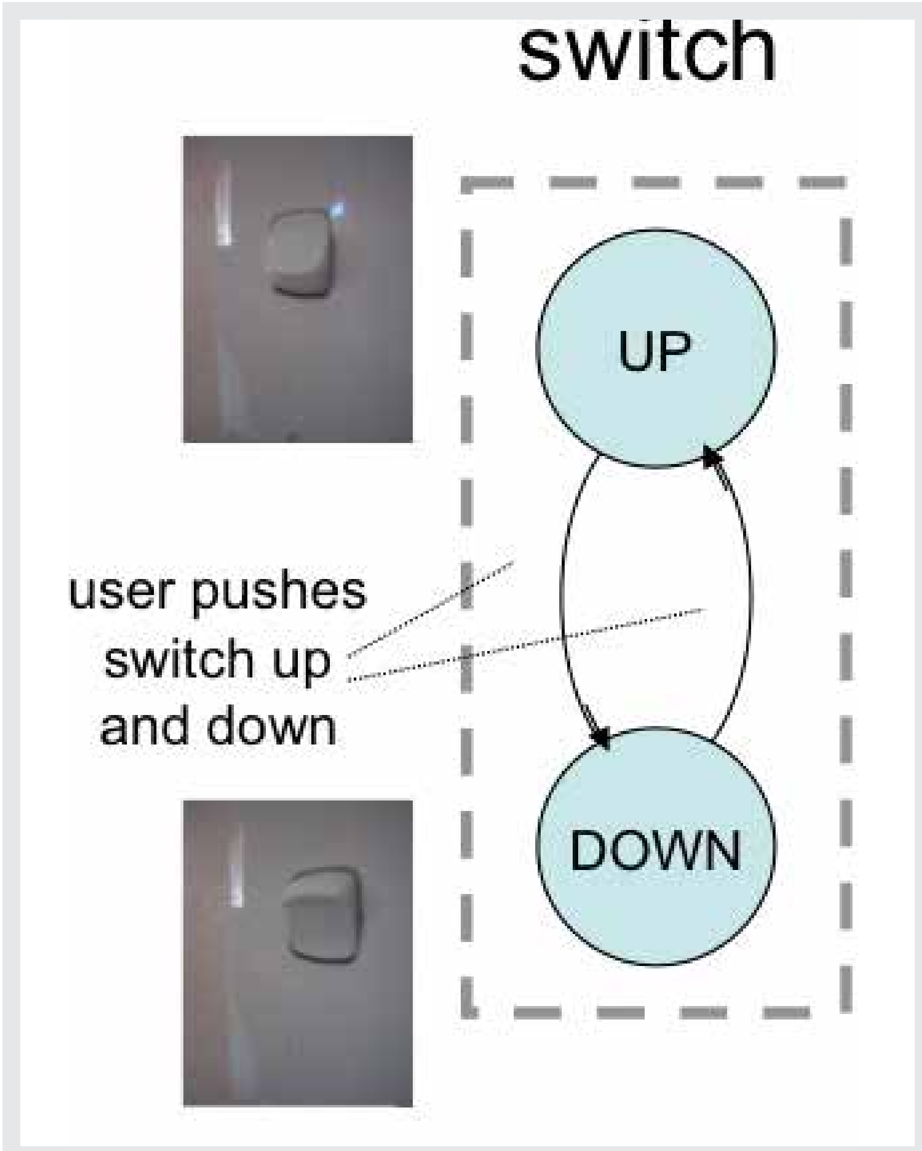
Figure 29.7: Deus Oculi - tableaux of beliefs as two people interact with the exhibit.



### 29.4.3 Physigrams - modelling physical aspects of digital devices

The humble light-switch is nearly as ubiquitous as the woord processor in the HCI literature. It has been analysed in terms of the cultural understanding implicit in its use, the mapping of switch direction to function (is down off or on?), the mapping between banks of switches and lights in a room, even accidentally switching off the lights mistaking it for the lift-call button (and the author once set off a fire alarm making the opposite mistake!). Here we are going to focus on the physical interactions of the light switch 'unplugged'.

We can think of the light switch as a two state device: lights on/off, you press the switch down and the light goes on, you lift it up and the lights go out. However, notice the "and the" in the previous sentence: there is first a physical interaction with the switch and then this gives rise to an effect on the light.


switch
UP
DOWN
user pushes
switch up
and down

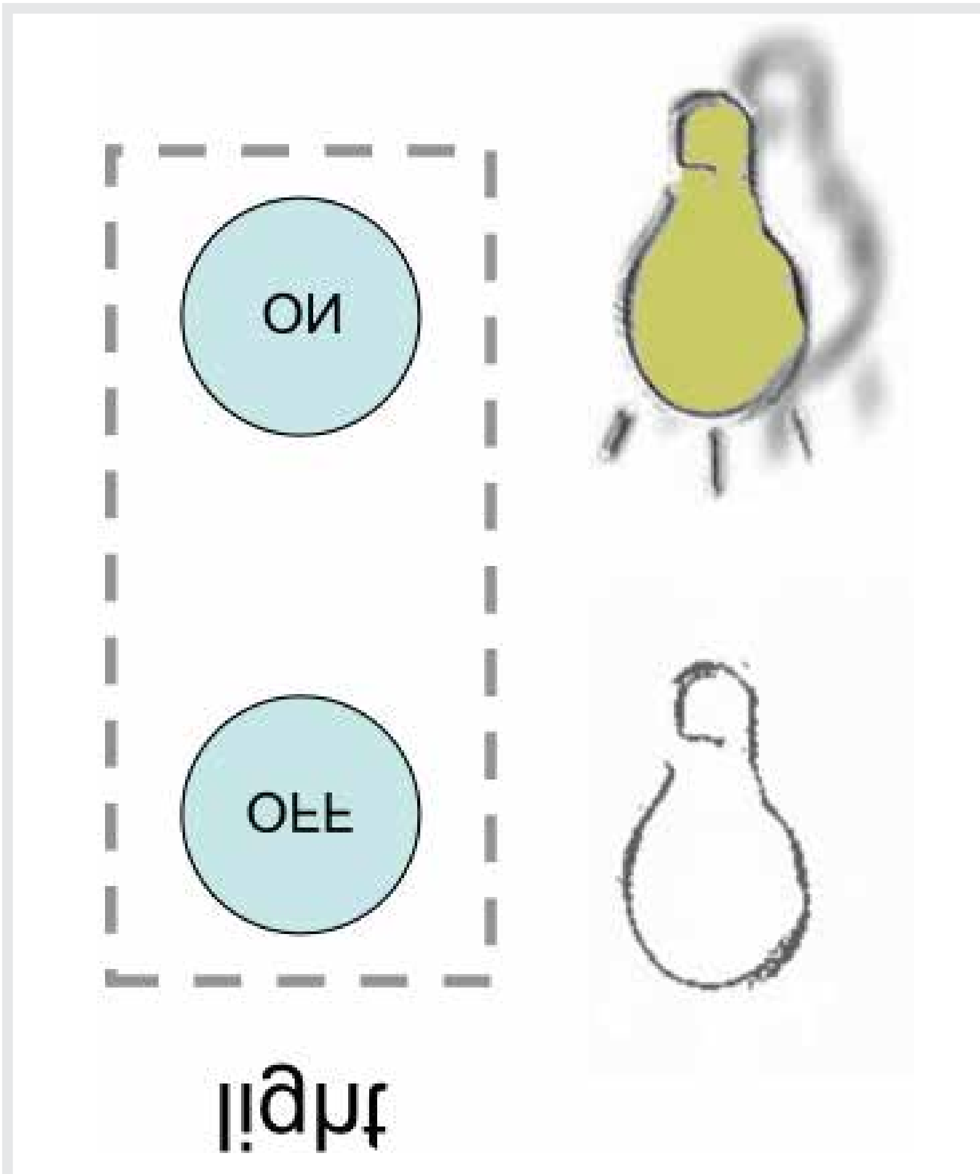





Figure 29.8 A-B: Light switch (A) physical device (B) logical states.

Figure 29.8 separates out these two. On the left (Figure 29.8.A) are the two states of the switch itself, on the right (Figure 29.8.B) the two states of the light. Drawing both allows us to discuss issues such as the mapping between switch position and light state. In this case, it is fairly obvious, but in more complex devices, this mapping may be something we wish to analyse at length. Crucial to any such discussion is having *both* diagrams.

The one on the right, the light on/off state, would often be the complex digital state of a device, for example the current channel selections on a television. This is the kind of thing that Formal Methods deal with frequently in software engineering, and typical of the level of detail used in Formal Methods in HCI.

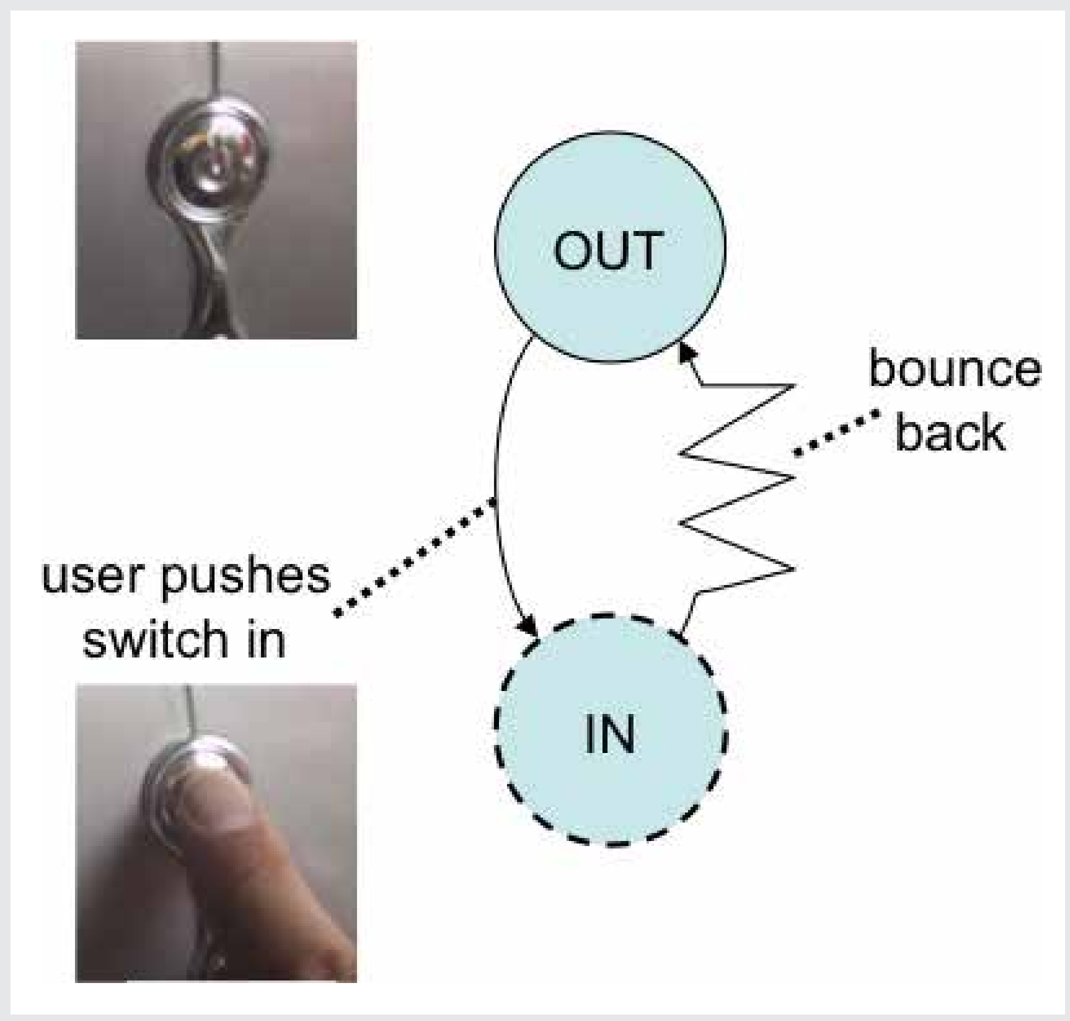


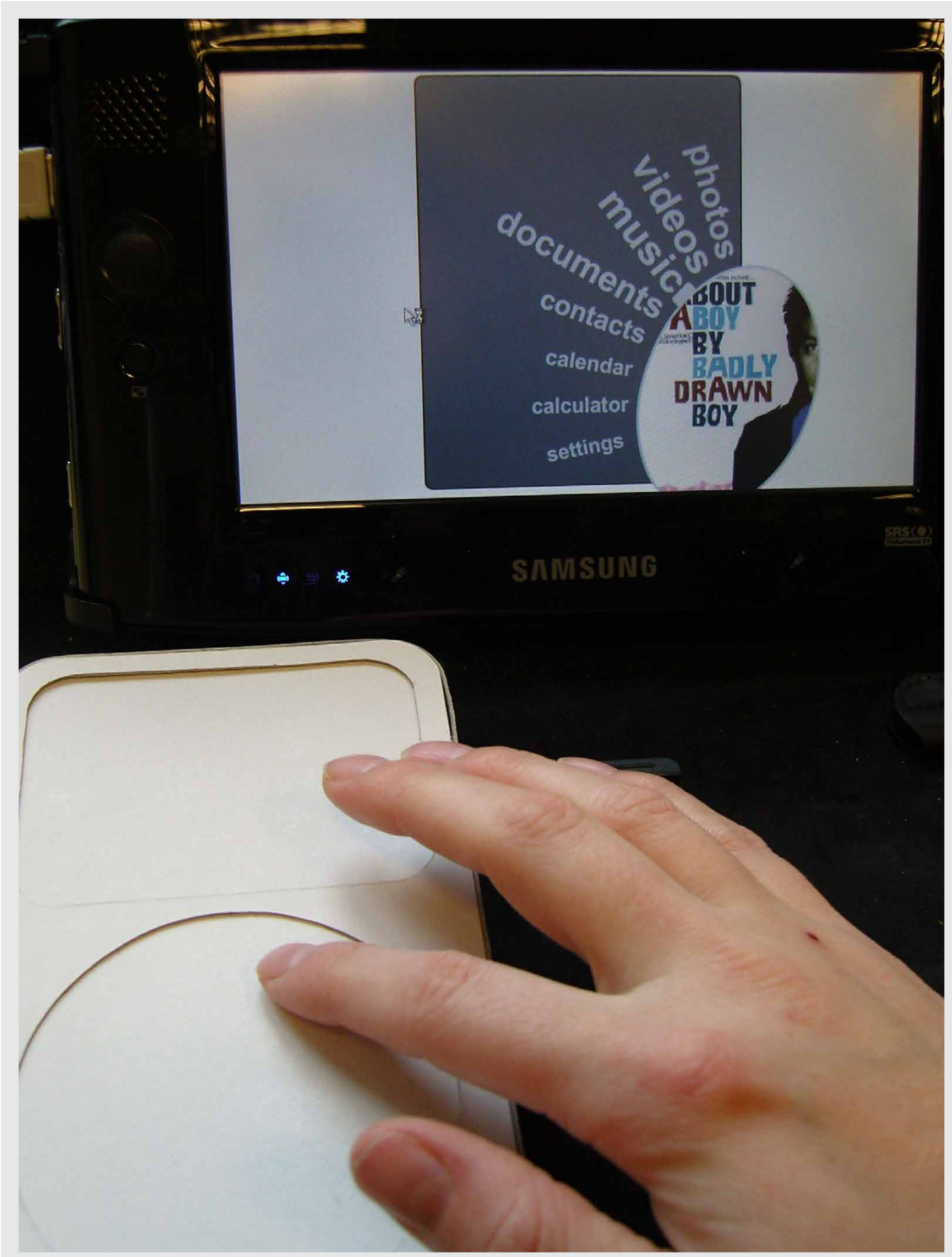

Figure 29.9: Computer light switch that 'bounces back' after you have pressed it.



In contrast, the one on the left is about the physical light switch itself. It captures the interaction properties/affordances inherent in its physical potential, even if it were torn from the wall and disconnected from electricity – literally 'unplugged'. Imagine a small child playing with an old switch, simply pushing it back and forth.

Physigrams are a modified form of state transition network (STN) designed to specify these physical interactions. They are like a normal STN, but have some additional aspects in order to deal with the fact that physical behaviours are different from digital ones. For example, Figure 29.9 shows the physigram of the kind of press button switch you often find on a computer. Unlike the simple light switch, it is almost always in the same position; it appears very much like a one-state device! However, in fact it does have an 'in' state, but it only stays there while you are pressing, a *tension* state. The dashed line around the 'in' state indicates that it is a tension state, and the lightning-bolt arrow denotes the fact that the transition occurs autonomously because of the internal spring within the switch.

Often Formal Methods are regarded as being too difficult for use by 'ordinary' designers, and only suitable for experts. However, this kind of diagrammatic representation is quite accessible (indeed you find STNs in consumer guides). Figure 29.10 A-B and 29.11 A-B shows physigrams produced by product designers using them to compare the properties of two physical device designs for the same underlying digital functionality. While the devices were similar in many ways (circular control, managing a scrolling menu), the physigrams highlighted key differences, particularly the lack of tactile feedback in one of the devices (scroll pad).


photos
videos
music
documents
contacts
calendar
calculator
settings
ABOUT A BOY BY BADLY DRAWN BOY
SAMSUNG

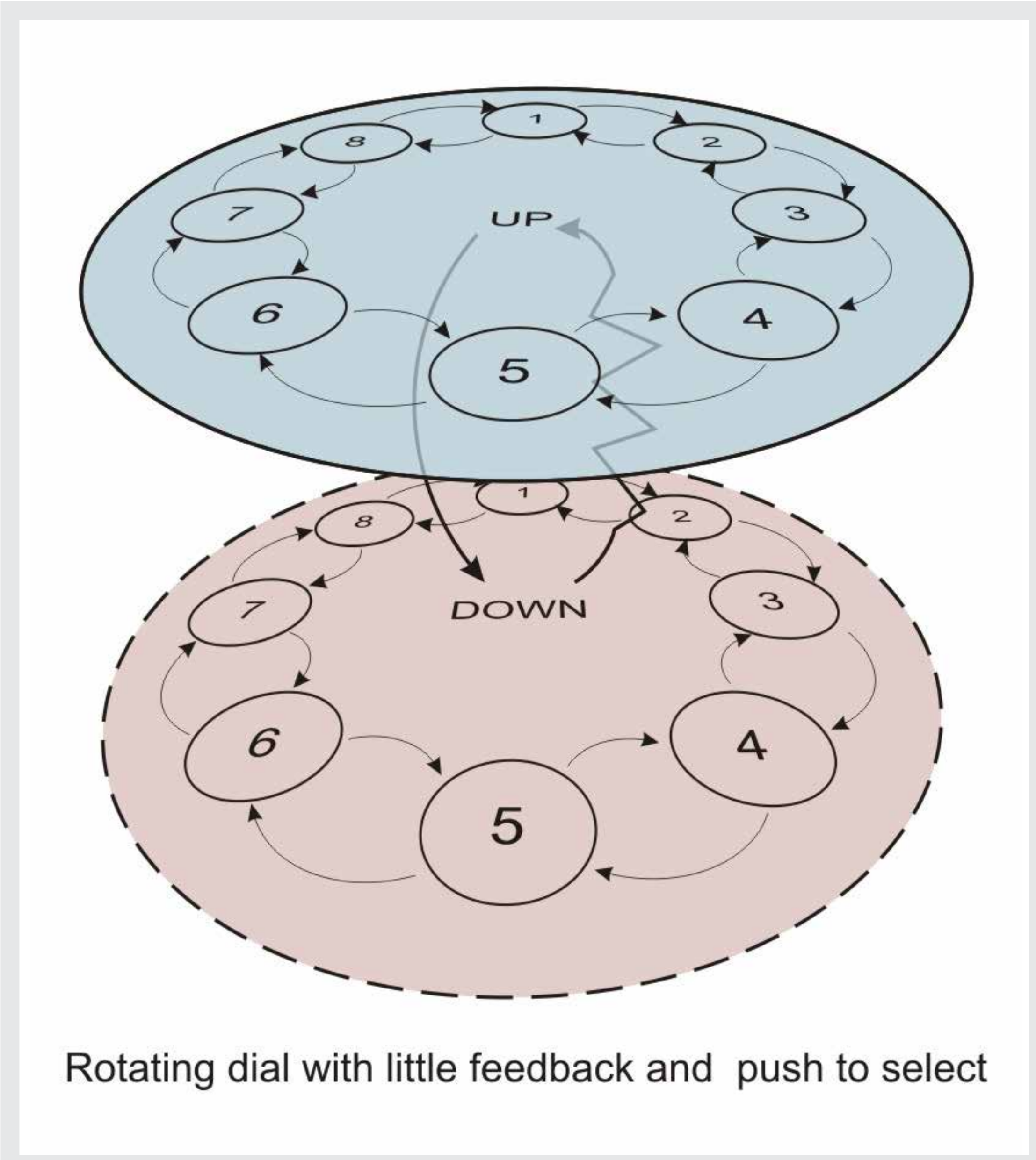





FIGURE 29.10 A-B: Physigrams created by product designers.

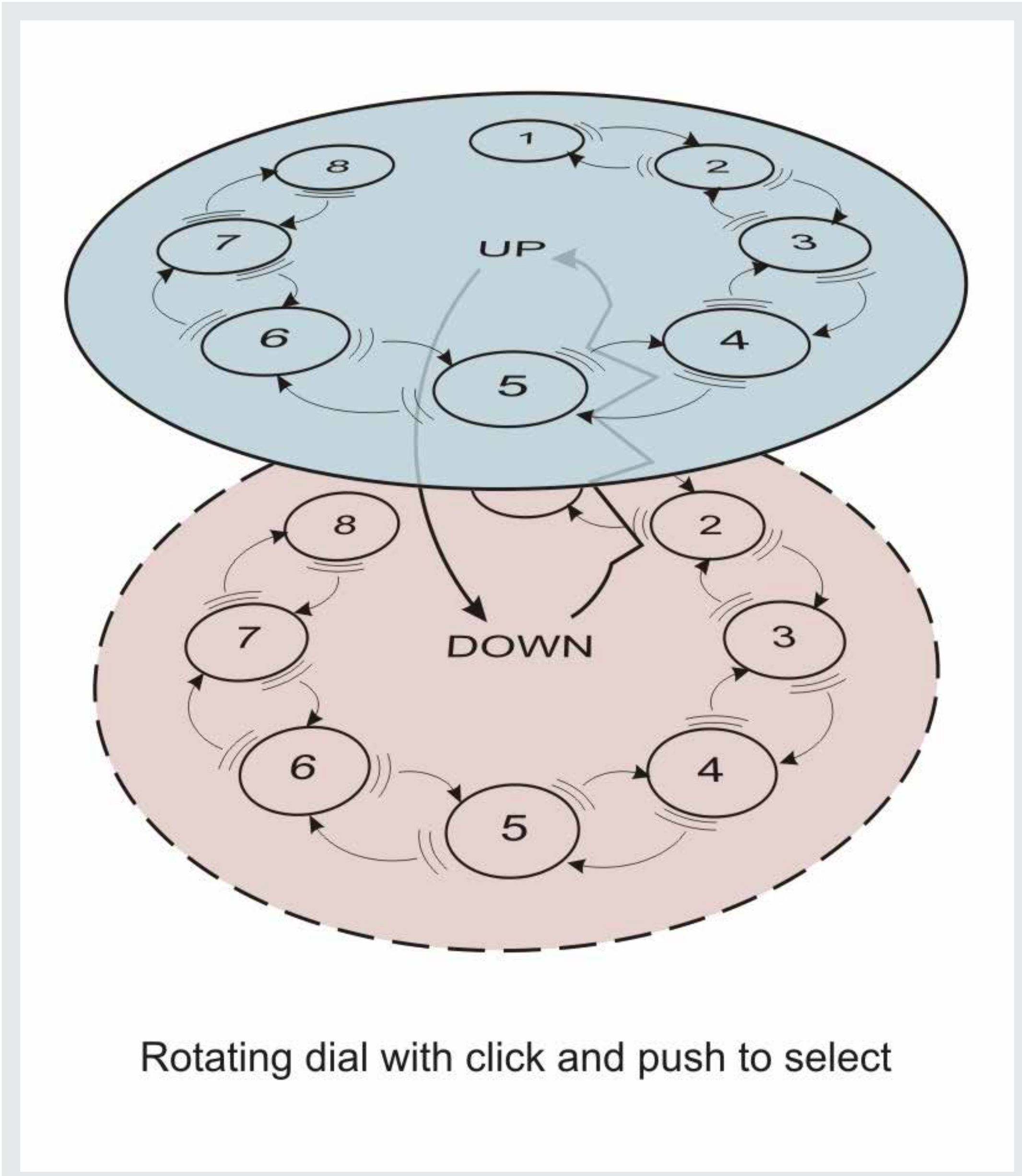





Figure 29.11 A-B: Physigrams created by product designers.

## 29.5 A SUCCESS STORY

It is often hard to measure the effectiveness of methods, especially if the claim is that they improve intangibles such as quality or maintainability. Furthermore, given the variability in people and projects, even large differences in measurable indicators, 30-50%, may be hard to detect. So, as an example, I am going to go back nearly 30 years, not because there haven't been any successes since then (!), but because the impact in this case was so dramatic, a 1000% (yes the ten-fold) increase in productivity.

### 29.5.1 The problem

This comes from personal experience, before I was an academic or had even heard of HCI. I was working for Cumbria County Council working on transaction processing programs in COBOL, the sort of thing used for stock control or point-of-sale terminals in large organisations.

While this all sounds like very old technology (mainframes!), in fact transaction processing systems are very like web-based systems: a message comes in from a terminal somewhere, it needs to be processed, a response is sent back, and the next message dealt with, usually from a completely different terminal.

The problems were not unlike those you sometimes see in web-based interfaces today: applications sometimes 'losing track' of where they are in interaction sequences, and exhibiting odd bugs, including once when a person on one terminal pressed 'next' on a listing screen and got an unexpected result, which turned out to be the 'next' page for a completely different terminal. Further more it took people a long time to produce this buggy code.

However, when you looked at the code, the lack of quality and efficiency was not surprising. Figure 29.12 gives a taste of what the code looked like.

```
if confirm_field is empty // can’t be confirm screen
 // or user didn’t fill in the Y/N box
then if record_id is empty // must be initial entry
 then prepare ‘which record to delete’ screen
 else if valid record_id
 then read record and prepare confirm screen
 else prepare error screen
else if confirm_field = “Y’
 then if record_id is empty // help malformed
 then prepare error screen
 else if valid record_id
 else do deletion
 then prepare error screen
 else if confirm_field = “N’
 then prepare ‘return to main menu’ screen
 else prepare ‘must answer Y/N’ screen
```

Figure 29.12: Typical 1980s transaction processing code.

The URL structure of the web makes this a little cleaner, and I know some wonderful web programmers creating beautifully clean code. However, if you have looked at a lot of web interface code, not least those from big open source projects, this nest of choices and sub-choices is not uncommon. The focus is on a single page, a single transaction, and the developer is in a “what just happened, what happens next?” mode of thinking.

### 29.5.2 The solution

Faced with this, I knew I would not be able to make sense of it and so looked for a way to make it manageable. The most obvious tool to hand was the humble flow chart. However, rather than create a flow chart of the program code (the normal use), I used flow charts such as the one in Figure 29.11 that mapped out the main screens and actions between them. That is, I created a dialogue specification. The boxes describing the computer actions between screens were typically described quite loosely, saving a record to disk might require a whole raft of shuffling variables, calling file system code, etc., but this is what programmers do. The flow chart was about what was difficult: keeping track of the big picture.

Sadly, there were no automated tools, but by hand the paper flow chart was translated into templated code. Each screen corresponded to a block of code that assembled the message to send to the terminal, and each block of computation code to run after the screen. The small identifiers were used as labels (goto!) in the code, so that when there was any problem, or any change needed to be made, it was possible to easily find the corresponding code. They were also put at the top corner of the screen as an identifier so that users could give them as a code if there were any problems to report. Finally, the flow diagram meant it was easy to see whether one had tested all parts through the dialogue (not the same as all oaths through the code).

The most important thing, though, was speed. It was much faster to draw the flow chart and translate it into code than it was to simply write the code (and this is without any support tools). Even more dramatic was the differences in turn-around time after discussions with users. It is important to recall that these were the days before rapid application development; typically it would take many days or weeks to create and update this type of applications, but instead changes could be implemented in hours, around 10 times faster than standard methods.

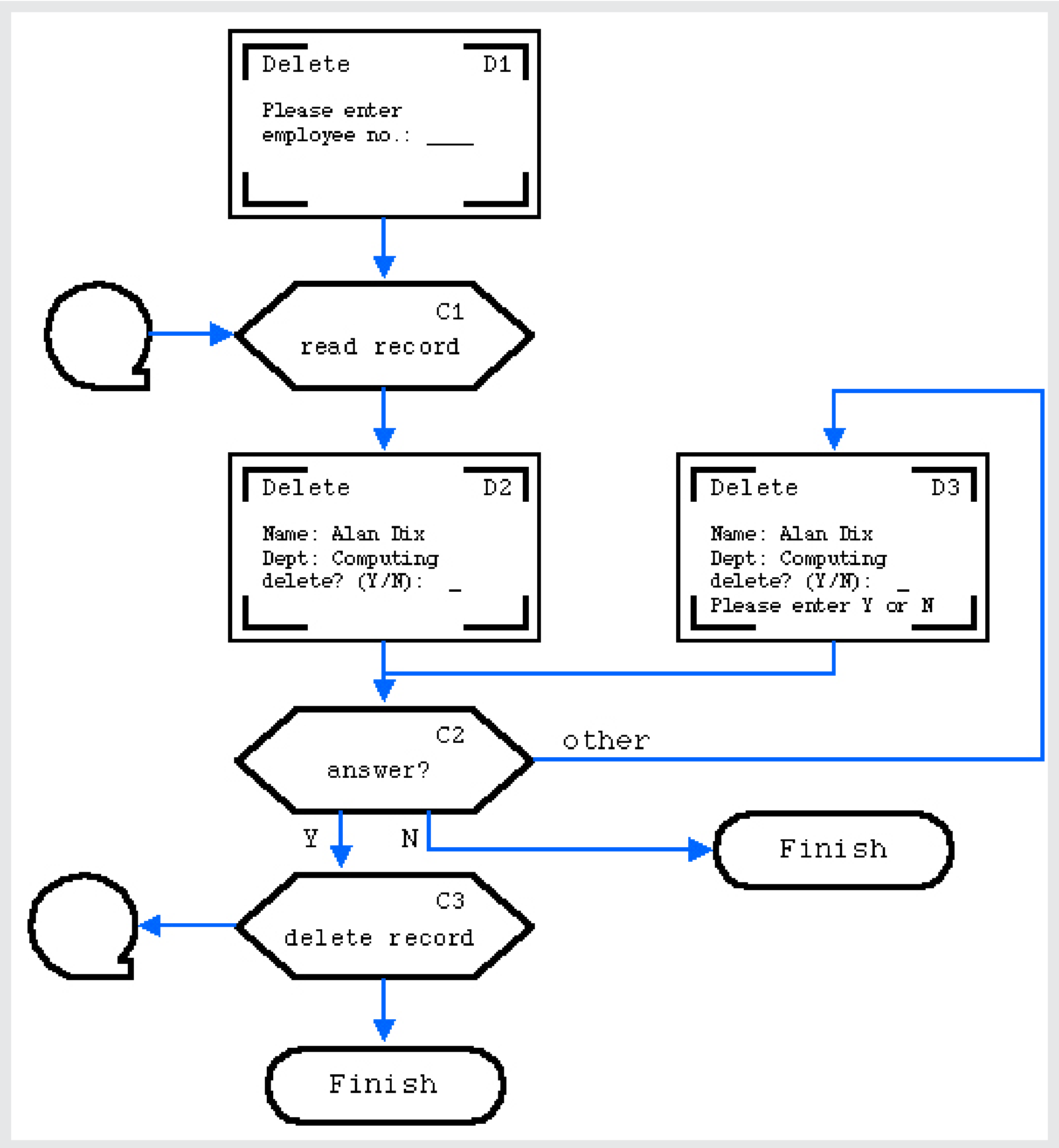



FIGURE 29.13: Flow chart of user interaction.

### 29.5.3 Why it worked?

While this was a very successful application of Formal Methods in interface design, I didn't understand *why* at the time. It is only by reflecting in the intervening years that I came to understand the rich interplay of factors that made it work and guidelines to help reproduce that success.

**useful** – addresses a real problem!
The notation focused on the overall user-interface dialogue structure that was causing difficulties in the existing systems. So, often formalisms are proposed because they have some nice intrinsic properties, or are good for something else, but do not solve a real need.

**appropriate** – no more detailed than needed
For example, there was no problem in producing the detailed code to access databases etc., so the formalism deals with this at a very crude level 'read record', 'delete record' etc. Many formalisms force you to fill in unnecessary detail, which makes it hard to see the things you really need it for as well as increasing the cost of using it.

**communication** – mini-pictures and clear flow easy to talk through with client
Formal Methods are often claimed to be a means to improve communication within a design team because of their precision. However, when precision is achieved at the cost of comprehensibility, there is no real communication.

**complementary** – different paradigm than implementation
It is common to use specification methods that reflect the final structure of the system – for example, object-oriented specification for object-oriented systems. Here, however, this specification represents the structure of the dialogue, which is completely different from the structure of the code. This is deliberate: the no-

tation allows one to see the system from a *different perspective*, in this case one more suitable for producing and assessing the interface design. The relationship between the structure of the notation and the structure of the code is managed via simple rules, which is what formalisms are good at!

**fast pay back** – quicker to produce application (at least 1000%)
The argument is often that Formal Methods will yield *long-term* time-savings. However, most people like instant pay-back. Spending lots of time up-front for savings later is very laudable, but developers and clients are often happier to see a buggy prototype than to be told that, "…yes, in a few months it will all come together." The dialogue flowcharts didn't just produce long-term savings, but also reduced the lead-time to see the first running system.

**responsive** – rapid turnaround of changes
The feeling of control and comprehension made it easier to safely make changes. In some Formal Methods, the transformation process between specification and code is so complex that change is very costly (see (Dix and Harrison 1989) for a discussion of this). Of course, with user interfaces, however well specified, it is only when they are used that we really come to fully understand the requirements.

**reliability** –clear boilerplate code less error-prone
Although the transformation process from diagram to code was not automated, it was a fairly automatic hand process applying and modifying boilerplate code templates. This heavy reuse of standard code fragments greatly increases the reliability of code.

**quality** –easy to establish test cycle
The clear labelling of diagrams and code made it easy to be able to track whether all paths had been tested. However, note that these are not just paths through the program (which effectively restarted at each transaction), but each path through the human–computer dialogue.

**maintenance** –easy to relate bug/enhancement reports to specification and code The screens presented to the user included the labels, making it easy to track bug reports or requests for changes both in the code and specification.

In short, the formalism was used to fulfil a purpose, and was, above all, neither precious nor purist!!

## 29.6 BARRIERS TO ACCEPTANCE

As noted in the introduction, Formal Methods have always sat a little uneasily within HCI, have never been popular with the CHI community, and can engender vitriol that is hard to explain in a discipline that is otherwise accepting.

A few years ago, as a meta-reviewer at an HCI conference I was forced to disregard one review completely as the reviewer's sole critique was to dismiss the paper simply because it was formal and this was clearly not an acceptable area; the author, it was suggested, should turn to ethnography or social science. Note, this was despite the fact that one of the workshops at the conference was on Formal Methods. It would be unthinkable to imagine a similar review dismissing an ethnographic or experimental paper simply because of the area.

To be fair, this probably reflects general trends in society (or at least Anglo-American society) where it is common to hear educated people boast about their innumeracy, whereas they would be deeply embarrassed by illiteracy. Furthermore, even within many computer science departments, the teaching of more formal topics has been reduced, as students' mathematical skills on entry do not make this, or topics such as graphics, achievable.

So, championing Formal Methods in HCI is not going to make you popular!

As well as requiring expertise, Formal Methods tend to be costly in terms of time taken, and it can be hard to justify large amounts of analysis time before any 'real' work appears to be done. There are long-term gains, it can be argued, but this is difficult to justify when the project is 3 months down the line and there is no code written yet. This has always been a problem, and more so with agile development techniques.

These cost barriers work at an individual area also. In my personal experience, when I have used Formal Methods extensively in real development, it has been obvious afterwards that the resulting systems were better designed, more robust and took less time to develop than more ad hoc approaches. However, despite knowing this, when faced with a new project, I just get in and hack! Extensive planning and preparation is worthwhile, but the bang for the buck comes late – it does not sit well with human psychology.

## 29.7 FUTURE DIRECTIONS

Despite the barriers above and general disopprobrium, there are a number of groups who continue to produce strong work in the area, often working closely with the industry, and the model-based user interface development is an active stream in W3C. Furthermore, the various venues for this work: the DSVIS, EHCI and TAMODIA conference series, have become part of the new jointly IFIP and ACM sponsored conference EICS (Engineering Interactive Computer Systems). This covers a range of areas, but Formal Methods is a strong strand.

Given the costs of applying Formal Methods, it is not supporting that the areas where they have been applied most consistently over the years have been in safety critical situations, where the impact of mistakes is high and so the costs of employing Formal Methods acceptable. This is the case with, for example, the group at Toulouse, who use variants of Petri Nets in their work with air traffic control, and the group at Swansea using graph methods with medical instruments.

Looking more generally, however, it seems the time is ripe for growth.

On the web interfaces have to coordinate multiple API services, have to target different devices, and have multiple ‘threads’ of execution (session, window, strict chronology). In our homes and workplaces, multiple devices are beginning to ‘talk’ to one another, so that our interactions with a phone might have an effect on our TV or washing-machine, our central heating is responding to manual controls, external temperature, and smart grid dynamic pricing. Traditional functional design broke things, including user interface elements, into intellectually manageable chunks. In contrast, the complex interactions between simple elements are hard to make sense of and can lead to emergent effects that do not show up during normal testing. In Agile development, automated testing is central to the discipline, and yet often the user interface is least well represented in the test suite.

These are issues crying out for Formal Methods!

Given this, the adoption of model-based techniques by W3C groups focused on adaptation to different devices is not surprising. Also in the UK, there are several large projects looking at the complex interactions of human agents and software systems on the web, and strands of Web Science clearly overlap with this.

At a practical level, it is interesting to observe practitioners creating ad hoc declarative representations at different levels within web-based systems; often using JSON, turtle, XML or other generic data-representations to create special purpose notations. Just as in the 'success story', they are using these formal representations (although they would not call them this) because they work, because they make their life easier. They are not a chore, not medicine that will eventually make you better, but of value now.

A weakness in the Formal Methods community has been to recognise that designers and developers, as well as users, are actually human. To be fair, this is also a charge that could be levelled at user-centred design in general, as all HCI methods tend to be up-front heavy with the promise of future rewards, hence the important job that Randolph Bias and Deborah Mayhew did in "Cost Justifying Usability" (Bias and Mayhew 1994).

I would love to see this addressed by the next generation of researchers (or the current generations!). To some extent, this is happening in the model-based development area, but the notations still feel top-heavy. Ideally, we would be creating notations and tools that require just enough effort at each stage of use to create value at that stage. However, the cumulative overall effect of adding this 'just enough' formality will mean that there is a sufficiently complete formal representation to use for further analysis.

What are the notations that would mean that those practitioners did not need to create their own ad hoc notations each time?

## 29.8 WHERE TO LEARN MORE

### 29.8.1 Books (in print or available online)

- A. J. Dix (1991). *Formal Methods for Interactive Systems*. Academic Press. ISBN 0-12-218315-0 http://www.hiraeth.com/books/formal/
  My own early monograph in the area. It includes PIE model and other models and methods. Although it is now 20 years old, I have recently found myself revisiting several of the techniques to apply to recent development projects.
  The book is out of print, but the full text is available (free!) as PDF on the book web site (above).
- Paternó, F. (2000). *Model-Based Design and Evaluation of Interactive Applications*. London, Springer-Verlag.
  Reviews a variety of models used in HCI, and then considers the whole design lifecycle using ConcurTaskTrees as a central representation. Note that the term 'model-based' here means models in general, not 'model-based user interface development' as discussed in Section 29.3.2, (although Fabio Paternó is also heavily involved in this area).
- H. Thimbleby, Press On — Principles of Interaction Programming, MIT Press, 2007, ISBN: 0262201704

Paperback edition, ISBN 978–0–262–51423–1, 2010.
This book is about programming interactive systems, but in it Harold explores in detail the use of various formalisms during construction and analysis,

### 29.8.2 Chapters

The following two chapters are available online, although sadly the excellent collections they were in are now out of print. Several sections of this encyclopaedia entry draw heavily on "Upside down As"

- A. J. Dix (1995). Formal Methods. In Perspectives on HCI: Diverse Approaches, Eds. A. Monk and N. Gilbert. London, Academic Press. pp. 9-43. http://www.alandix.com/academic/papers/formal-chapter-95/
- Upside down As and algorithms - computational formalisms and theory In HCI Models, Theories, and Frameworks: Toward an Multidisciplinary Science. John Carroll (ed.) ISBN 1-55860-808-7. Morgan Kaufman, 2003. pp. 381-429 http://www.alandix.com/academic/papers/theory-formal-2003/
- Two chapters from my HCI textbook deal extensively with issues connected with formalism:
- Chapter 16: Dialogue notations and design
  Chapter 17: Models of the system
  A. Dix, J. Finlay, G. Abowd and R. Beale (2004). *Human-Computer Interaction, third edition*. Prentice Hall. ISBN 0-13-239864-8.

### 29.8.3 Conferences

**EICS: ACM SIGCHI Symposium on Engineering Interactive Computer Systems** (annual since 2009). eics-conference.org This has a broader remit than purely Formal Methods, but the main players in the area tend to report there. Note that a number of older conferences amalgamated to form EICS (see below).

**FMIS: Formal Methods for Interactive Systems** (annual/biannual since 2006) fmis.iist.unu.edu/fmis_events.html Smaller workshop dedicated solely to the topic.

### 29.8.4 Conferences (older)

The following conferences have amalgamated into EICS, but many key papers can be found in the old proceedings.

- **DSVIS: Design, specification, and verification of interactive systems.** (annual 1994–2008). Probably the key annual venue for Formal Methods in HCI (although not limited to this).
- **EHCI: Engineering Human–Computer Interaction (triennial).** Periodic conference organised by IFIP WG2.7/13.4 (User Interface Engineering), which is the only IFIP working group cross listed under two technical committees.
- **TAMODIA: Task Models and Diagrams** (annual, 2002-2010)
- **CADUI: Computer Aided Design of User Interfaces** (annual 2002??–2008)

### 29.8.5 Books (not in print, but maybe in library)

These are out of print, but a few copies are available second-hand online.

- Harrison, M.D. and Thimbleby, H.W., editors (1990). Formal Methods in Human Computer Interaction. Cambridge: Cambridge University Press.
  Harrison and Thimbleby's 1990 collection includes contributions

from many of those working in the early days of this area.
My own chapter from this is available online at: http://www.alan-dix.com/academic/papers/nond90/

- Palanque, P. and Paternó, F., editors (1997). Formal Methods in Human Computer Interaction. London, Springer-Verlag.
  Palanque and Paternó edited a more recent collection, confusingly with the same name as the one above! This collection is thematic with the contributors using their various techniques to address the web browser as a common example.
  My own chapter from this is available online at: http://www.alan-dix.com/academic/papers/histchap97/

## 29.9 REFERENCES


*Anderson, John R. (2005): Human Symbol Manipulation Within an Integrated Cognitive Architecture. In Cognitive Science, 29 (3) pp. 313-341*

*Bias, Randolph G. and Mayhew, Deborah J. (eds.) (1994): Cost-Justifying Usability. Boston, MA, Academic Press*

*Calvary et al. 2002. Calvary, G., Coutaz, J., Bouillon, L., Florins, M., Limbourg, Q., Marucci, L., Paternò, F., Santoro, C., Souchon, N., Thevenin, D., Vanderdonckt, J., 2002 The CAMELEON Reference Framework, Deliverable 1.1, CAMELEON Project. http://giove.isti.cnr.it/projects/cameleon/deliverable1_1.htm*

*Campos, J. C. and Harrison, Michael D. (2001): Model checking interactor specifications. In Automated Software Engineering, 8 (3) pp. 275-310*

*Card, Stuart K., Moran, Thomas P. and Newell, Allen (1983): The Psychology of Human-Computer Interaction. Hillsdale, NJ, Lawrence Erlbaum Associates*

*Card, Stuart K., Moran, Thomas P. and Newell, Allen (1980): The keystroke-level model for user performance with interactive systems. In Communications of the ACM, 23 pp. 396-410*

*Diaper, Dan and Stanton, Neville (2003): The Handbook of Task Analysis for HCI. Hillsdale, New Jersey, USA, Lawrence Erlbaum Associates*

*Dix, Alan J. (2003): Upside-down A's and Algorithms - Computational Formalisms and*

*Theory. In: Carroll, John M. (ed.). "HCI Models, Theories, and Frameworks". San Francisco: Morgan Kaufman Publisherspp. 381-428*

Dix, Alan J. (1991): *Formal Methods For Interactive Systems. Academic Press*

Dix, Alan J. (2011). *Are five users enough? HCI Book: Online. Retrieved 31 January 2012 from http://www.hcibook.com/e3/online/are-five-users-enough/*

Dix, Alan J. (1987): *The Myth of the Infinitely Fast Machine.. In: Diaper, Dan and Winder,* Russel (eds.) *Proceedings of the Third Conference of the British Computer Society Human Computer Interaction Specialist Group - People and Computers III* August 7-11, 1987, University of Exeter, UK. pp. 215-228

Dix, Alan J. (1990): *Non-determinism as a paradigm for understanding the user interface. In: Harrison,* Michael D. and Thimbleby, Harold (eds.). *"Formal Methods in Human-Computer Interaction (Cambridge Series on Human-Computer Interaction)". Cambridge University Press*

Dix, Alan J. and Harrison, Michael D. (1989): *Interactive systems design and formal development are incompatible?. In: McDermid, John A. (ed.). "The Theory and Practice of Refinement: Approaches to the Formal Development of Large-Scale Software Systems". Butterworth-Heinemannpp. 12-26*

Dix, Alan J. and Runciman, Colin (1985): *Abstract Models of Interactive Systems. In: Johnson,* Peter and Cook, Stephen (eds.) *Proceedings of the Conference of the British Computer Society Human Computer Interaction Specialist Group - People and Computers I* August 17-20, 1985, University of East Anglia. pp. 13-22

Dix, Alan J., Mancini, Roberta and Levialdi, Stefano (1997): *The Cube - Extending Systems for Undo. In: In Proc. of DSVIS97* 1997.

Dix, Alan J., Rodden, Tom, Davies, Nigel, Trevor, Jonathan, Friday, Adrian and Palfreyman, Kevin (2000): *Exploiting Space and Location as a Design Framework for Interactive Mobile Systems. In ACM Transactions on Computer-Human Interaction, 7 (3) pp. 285-321*

Dix, Alan J., Finlay, Janet E., Abowd, Gregory D. and Beale, Russell (2004): *Human-Computer Interaction (3rd Edition). Prentice Hall*

Dix, Alan J., Sheridan, Jennifer G., Reeves, Stuart, Benford, Steve and O'Malley, Claire (2005): *Formalising Performative Interaction. In: Gilroy,* Stephen W. and Harrison, Michael D. (eds.) *DSV-IS 2005 - Interactive Systems, Design, Specification, and Verification, 12th International Workshop* July 13-15, 2005, Newcastle upon Tyne, UK. pp. 15-25

*Ellis, Clarence A. (1994): Goal Based Workflow Systems. In International Journal of Collaborative Computing, 1 (1) pp. 61-86*

*Eslambolchilar, Parisa (2006). Making Sense of Interaction Using a Model-Based Approach. PhD Thesis. Hamilton Institute, National University of Ireland*

*Eslambolchilar, Parisa and Murray-Smith, Roderick (2010): A Model-Based Approach to Analysis and Calibration of Sensor-Based Human Interaction Loops. In International Journal of Mobile Human Computer Interaction, 2 (1) pp. 48-72*

*Hillier, Bill (1999): Space is the Machine: A Configurational Theory of Architecture. Cambridge University Press*

*Kostakos, Vassilis, O'Neill, Eamonn, Penn, Alan, Roussos, George and Papadongonas, Dikaios (2010): Brief encounters: Sensing, modeling and visualizing urban mobility and copresence networks. In ACM Transactions on Computer-Human Interaction, 17 (1) p. 2*

*Laird, John E. (2008): Extending the Soar Cognitive Architecture. In: Proceedings of the 2008 conference on Artificial General Intelligence 2008 Proceedings of the First AGI Conference 2008. pp. 224-235*

*Larson, Kevin and Czerwinski, Mary (1998): Web Page Design: Implications of Memory, Structure and Scent for Information Retrieval. In: Karat, Clare-Marie, Lund, Arnold, Coutaz, Joëlle and Karat, John (eds.) Proceedings of the ACM CHI 98 Human Factors in Computing Systems Conference April 18-23, 1998, Los Angeles, California. pp. 25-32*

*Limbourg, Quentin, Vanderdonckt, Jean M., Michotte, Benjamin, Bouillon, Laurent and López-Jaquero, Víctor (2005): USIXML: A Language Supporting Multi-path Development of User Interfaces. In: Bastide, Remi, Palanque, Philippe A. and Roth, Jörg (eds.) Engineering Human Computer Interaction and Interactive Systems, Joint Working Conferences EHCI-DSVIS 2004 July 11-13, 2005, Hamburg, Germany. pp. 200-220*

*Lin, James, Newman, Mark W., Hong, Jason I. and Landay, James A. (2000): DENIM: Finding a Tighter Fit between Tools and Practice for Web Site Design. In: Turner, Thea, Szwillus, Gerd, Czerwinski, Mary, Peterno, Fabio and Pemberton, Steven (eds.) Proceedings of the ACM CHI 2000 Human Factors in Computing Systems Conference April 1-6, 2000, The Hague, The Netherlands. pp. 510-517*

*Lynch, Kevin (1960): The Image of the City (Harvard-MIT Joint Center for Urban Studies Series). The MIT Press*

*Miller, George A. (1956): The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information. In Psychological Review, 63 pp. 81-97*

Navarre, David, Palanque, Philippe A., Paterno, Fabio, Santoro, Carmen and Bastide, Remi (2001): A Tool Suite for Integrating Task and System Models through Scenarios. In: Johnson, Chris (ed.) DSV-IS 2001 - Interactive Systems Design, Specification, and Verification, 8th International Workshop June 13-15, 2001, Glasgow, Scotland, UK. pp. 88-113

Nielsen, Jakob and Landauer, Thomas K. (1993): A Mathematical Model of the Finding of Usability Problems. In: Ashlund, Stacey, Mullet, Kevin, Henderson, Austin, Hollnagel, Erik and White, Ted (eds.) Proceedings of the ACM CHI 93 Human Factors in Computing Systems Conference April 24-29, 1993, Amsterdam, The Netherlands. pp. 206-213

Paterno, Fabio (1999): Model-Based Design and Evaluation of Interactive Applications (Applied Computing). Springer

Pfaff, Günther E. (ed.) (1985): User Interface Management Systems (Eurographic Seminars). Springer

Pirolli, Peter (2007): Information Foraging Theory: Adaptive Interaction with Information (Series in Human-Technology Interaction). Oxford University Press, USA

Prinz, Wolfgang, Pankoke-Babatz, Uta, Gräther, Wolfgang, Gross, Tom, Kolvenbach, Sabine and Schäfer, Leonie (2004): Presenting Activity Information in an Inhabited Information Space. In: Snowdon, David N., Churchill, Elizabeth F. and Frécon, Emmanuel (eds.). "Inhabited Information Spaces: Living with your Data (Computer Supported Cooperative Work)". Springer

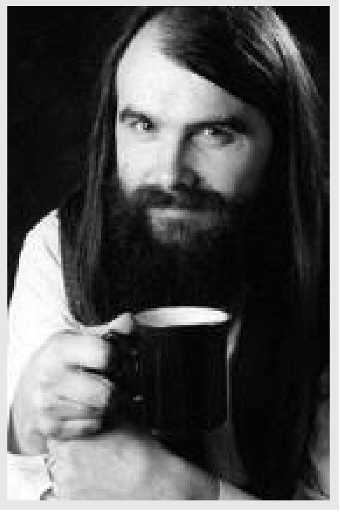
Reisner, Phyllis (1981): Formal Grammar and Human Factors Design of an Interactive Graphics System. In IEEE Trans. Software Eng., 7 (2) pp. 229-240

Thimbleby, Harold and Cairns, Paul (2010): Reducing number entry errors: solving a widespread, serious problem. In Journal of The Royal Society Interface, 7 (51) pp. 1429-1439

Université catholique de Louvain (2011). UsiXML - Home of the USer Interface eXtensible Markup Language. Retrieved 29 March 2011 from Université catholique de Louvain: http://www.usixml.org/

## About the author

### Alan J. Dix



© *Alan J. Dix*

Has also published under the name of:
"Alan John Dix"

Personal Homepage:
http://www.alandix.com/academic/

I am a computing professor at Lancaster University and researcher at Talis Ltd. and work on most things that connect people and computers. However, I started off (years ago) as mathematician and this is still my academic first love!

## Your notes and thoughts on chapter 29

Record your notes and thoughts on this chapter. If you want to share these thoughts with others online, go to the bottom of the page at: http://www.interaction-design.org/encyclopedia/formal_methods.html

**NOTES:**

# Chapter 30

# Personas

*by Lene Nielsen.*

The persona method has developed from being a method for IT system development to being used in many other contexts, including development of products, marketing, planning of communication, and service design. Despite the fact that the method has existed since the late 1990s, there is still no clear definition of what the method encompasses. Common understanding is that the persona is a description of a fictitious person, but whether this description is based on assumptions or data is not clear, and opinions also differ on what the persona description should cover. Furthermore, there is no agreement on the benefits of the method in the design process; the benefits are seen as ranging from increasing the focus on users and their needs, to being an effective communication tool, to having direct design influence, such as leading to better design decisions and defining the product's feature set (Cooper, 1999; Cooper et al, 2007; Grudin & Pruitt, 2002; Long, 2009; Ma & LeRouge, 2007; Miaskiewicz & Kozar, 2011; Pruitt & Adlin, 2006).

A persona is not the same as an archetype or a person. The special aspect of a persona description is that you do not look at the entire person, but use the area of focus or domain you are working within as a lens to highlight the relevant attitudes and the specific context associated with the area of work.

## 30.1 AN EXAMPLE PERSONA

Dorte is 53 years old and works as a secretary in her husband's plumbing business in the suburbs of Copenhagen. There are 5-6 assistants and apprentices in the company.

*Background*

When Dorte was very young she trained as an office clerk in the accounts department in a department store in Copenhagen. She was married at the age of 21 to Jan who had just got his skilled worker's certificate. They have two grown-up sons who no longer live at home in the combined house and workshop/office. Their sons visit frequently as they still enjoy mum's cooking.

Dorte likes to keep up with fashion. She often goes to the hairdresser, loves vibrant colours and elegant shoes. When she reads women's magazines, she looks for small tips that she changes and makes her own. She is always smartly dressed and stays fit.

Dorte loves travelling to faraway countries; most recently, she and her family were on a trip to Vietnam this summer. Before they went, she spent time reading up on the country and also watched the film Indochine starring Catherine Deneuve. Dorte always discusses the vacations with Jan, who would prefer to go to Rhodes with old friends, but it is Dorte who has the final say about the destination.

In an average day, she tends to drink too many cups of coffee, and when the telephone rings all the time and she can't reach the assistants, she also tends to smoke a bit too much.

Dorte makes payments to the Danish early retirement benefit scheme and looks forward to the day where she no longer has to be the "mum" of others any more and can spend more time travelling.

*Computer use*

Dorte does the accounts and the bookkeeping, VAT, taxes, vacation pay, the Danish Labour Market Supplementary Pension ATP, etc. She uses a mini financial management system that she has mastered after many years of use, but sometimes the system is not completely logical.

If she were to use other systems or use new, digital reporting, she would prefer it to be demonstrated to her by someone. She feels unable to learn something new when it is just explained to her, and she dislikes reading user guides. She says it takes her a long time to study anything new and familiarise herself with it, and she tends to see more limitations than possibilities in new IT. Dorte often underestimates her IT proficiency and overestimates the time that it will take to learn something new, so she stalls before she even gets started.

If she needs IT help, her oldest son and, less often, a woman friend provide the support. The friend works in a big company and is a super-user of the financial management software.

*Reporting*

Dorte handles the tax cards for the business. She deals with and reports the wages, vacations, sickness benefits, and maternity leaves of the staff. She does the VAT returns and annual accounts of the company. In addition, she fills in the reports for Statistics Denmark and the Employer's Reimbursement System AER.

Dorte does not understand the logic of the IT system and does not trust everything to happen as it should. If she sends in a return form or report digitally, she likes a confirmation saying that the recipient has received the form.

*Her workday:*

- She is not involved in the plumbing business as a trade, but she knows all the technical terms.
- She tidies things up. She does not want the others (her husband and the assistants) to make a mess in the basement where the office is as she is the one who has to look at it all day "Tidy up! Your mum does not work here!"
- She digs in and sometimes has to keep far too many balls in the air at the same time.
- She holds the fort, but does not get a lot of professional recognition in the company from the boss/her husband.

- She answers the telephone, handles mail, deliveries of goods (including invoices and delivery letters), and email.
- She handles the accounts, does some bookkeeping and writes invoices.
- She makes the coffee.
- She has occasional contact with the accountant.
- She does the invoicing of clients.
- She sends/delivers mail every day.
- She sends reminders.
- She handles customer contact (including damage control).
- She also walks the dog.

*Future goals*

Dorte dreams about a future where she no longer has to work and where she can spend more time travelling. She is still debating with Jan whether they should travel or buy a summer cottage where they can live all year round when they retire.

FIGURE 30.1: Persona for Virk.dk. Virk.dk is a portal for digital reporting. At Virk.dk, Danish companies can find all the forms needed for reporting to the authorities.

The persona approach stems from IT system development where in the late 1990s many researchers had begun reflecting on how you could communicate an understanding of the users. In literature, various concepts emerged, such as user archetypes, user models, lifestyle snapshots, and model users — as I termed it when I first wrote about the method in 1997. In 1999, Alan Cooper published his tremendously successful *The Inmates are Running the Asylum* (Cooper, 1999) where the persona as a concept to describe fictitious users was introduced for the first time. Despite the fact that a vast number of articles about using personas have been written, there is no unilateral understanding of the application of the method nor a definition of what a persona description is.

## 30.2 FOUR DIFFERENT PERSPECTIVES

The literature today offers four different perspectives regarding personas: Alan Cooper's goal-directed perspective; Jonathan Grudin, John Pruitt and Tamara Adlin's role-based perspective; the engaging perspective that I myself use, which emphasizes how the story can engage the reader (Sønderstrup-Andersen, 2007); and the fiction-based perspective. The first three perspectives agree that the persona descriptions should be founded on data. However, the fourth perspective, the fiction-based perspective, does not include data as the basis for persona description, but creates personas from the designers' intuition and assumptions; they have names such as ad hoc personas, assumption personas, and extreme characters.

### 30.2.1 The goal-directed perspective

Cooper characterizes his persona method as "Goal-Directed Design" and maintains that it makes the designer understand the user. Thus, Goal Directed Design is meant as an efficient psychological tool for looking at problems and a guide for the design process. The central core of the method is the hypothetical archetype that is not described as an average person, but rather as a unique character with specific details.

The method focuses on a move from initial personas to final personas. In the beginning of the process, a large number of personas are created based on in-depth ethnographic research. The initial personas grasp an intuitive understanding of user characteristics. Later on, these are condensed into final personas, one persona for each kind of user. Every project has its own set of personas (Floyd et al., 2008).

A persona is defined by its personal, practical, and company-oriented goals as well as by the relationship with the product to be designed, the emotions of the persona when using the product, and the goals of the persona in using the product (hence Goal-Directed).

In other words, it is the users' (work) goals that are the focus of the persona descriptions, e.g. workflow, contexts, and attitudes. And, as implied, the advantage of the method is that it provides a focused design and a communication tool to finish discussions.

### 30.2.2 The role-based perspective

The role-based perspective shares goal direction with Cooper and also focuses on behaviour. The personas of the role-based perspective are massively data-driven and incorporate data from both qualitative and quantitative sources. Often, dozens of personas are shared among projects (Floyd et al, 2008).

The starting point of the role-based perspective was criticism of the traditional IT system development approaches and of Cooper's approach to personas. The traditional use of scenarios is criticized for lacking clarity and consistency in the user descriptions. Therefore, the critics introduced user archetypes, which can communicate the most important knowledge about the users and thereby support the design process. Jonathan Grudin and John Pruitt criticized Alan Cooper for underestimating the value of user involvement and for seeing the method as one single method that can handle anything (Mikkelson & Lee, 2000),(Grudin & Pruitt, 2002).

The role-based perspective used the criticism as a starting point to develop the method further. The most important additions are, firstly that both qualitative and quantitative materials must supplement the persona descriptions; and secondly that there should be a clear relationship between data and the persona description (Grudin & Pruitt, 2002). Personas can communicate more than design decisions to designers and clients; they can also communicate information from market research, usability tests, and prototypes to all participants in the project. Finally, the method is regarded as a usability method that cannot stand alone, but should be used in tandem with other methods. The persona description itself should contain information about several issues: how big a share of the market the individual persona takes up; how much market influence the persona has; the user's computer proficiency, activities, and hopes and fears; and a description of a typical day or week in the life of the user. In addition to this are strategic and tactical considerations (Pruitt & Adlin, 2006).

The role-based perspective focuses on the users' roles in the organization (Sønderstrup-Andersen, 2007). Personas are an efficient design tool because of our cognitive ability to use fragmented and incomplete knowledge to form a complete vision of the people who surround us. With personas, this ability comes into play in the design process, and the advantage is that a greater sense of involvement and a better understanding of reality will be created.

### 30.2.3 The engaging perspective

The engaging perspective is rooted in the ability of stories to produce involvement and insight. Through an understanding of characters and stories, it is possible to create a vivid and realistic description of fictitious people. The purpose of the engaging perspective is to move from designers seeing the user as a stereotype with whom they are unable to identify and whose life they cannot envision, to designers actively involving themselves in the lives of the personas. The other persona perspectives are criticized for causing a risk of stereotypical descriptions by not looking at the whole person, but instead focusing only on behaviour (Nielsen, 2004;Nielsen, 2011; Nielsen 2012).

The starting point for the engaging perspective is the way we as humans interact with other people. We experience specific meetings in time and place. We mirror ourselves in the people we meet. And we experience others as both identical to and different from ourselves. Also, we experience relationships that are not specific and where the person we meet is anonymous and represents a type. Here, we use our experiences to understand the person and to predict what actions he or she will perform. If the designers see the users as stereotypical representations, they mould a mental image of the users together with a number of typical and automated acts. These representations prevent insight into the unique situation of the users and reduce the value of the scenario as a tool to investigate and describe future solutions.

An engaging description requires a broad knowledge of the users, and data should include information about the social backgrounds of the users, their psychological characteristics, and their emotional relationship with the focus area. The persona descriptions balance data and knowledge about real applications and fictitious information that is intended to evoke empathy. This way, the persona method is a defence against automated thinking.

### 30.2.4 The fiction-based perspective

The personas in the fiction-based perspective are often used to explore design and generate discussion and insights in the field (Floyd et al., 2008). Ad hoc personas are based on the designers' intuition and experience and used to create an empathetic focus in the design process (Norman, 2004). Extreme characters help to generate design insights and explore the edges of the design space (Djajadiningrat et al, 2000). Assumption personas are based on the project teams' assumed understanding of their users (Adlin & Pruitt, 2006). Proto-personas originate from brainstorming workshops, where company participants try to encapsulate the organization's beliefs (based on their domain expertise and gut feeling) about who is using their product or service and what is motivating them to do so. They give an organization a starting point from which to begin evaluating its products and to come up with some early design hypotheses (Gothelf, 2012). Pastiche scenarios create personas derived from fiction, like Bridget Jones or Ebenezer Scrooge, and help designers to be reflexive when creating scenarios (Blythe & Wright 2006). Examples of pastiche scenarios can be seen at Mark Blythe's website.

These personas have spurred discussions about validity and value (see e.g. When does a Persona stop being a Persona and Assumption personas help overcome hurdles). In line with Adlin & Pruitt (2006), James Robertson (2008) in his article *Beyond Fake Personas* suggests a continuum from Persona Sketch, over Persona Hypothesis and Provisional Personas, to Robust Personas ending in Complete Personas.

### 30.2.5 Other perspectives

Floyd et. al. 2008 mention three additional types that they have come across: Quantitative data driven personas are extracted from natural groupings in quantitative data: User archetypes as personas are similar to personas, but more generic, usually defined by role or position: Finally Marketing personas are created for marketing reasons and not to support design.

### 30.2.6 Criticism

Criticism of the persona method pertains to empiricism, especially the relationship between data and fiction. The implementation of the persona method in companies has also come under fire (Chapman et al, 2008; Chapman & Milham, 2006; Portigal, 2008; Rönkkö et al, 2004).

Because the persona descriptions have fictitious elements, some find it difficult to see the relationship with real users and the way that the data used is collected and analysed. Furthermore, the fictitious elements apparently prevent the method from being regarded as scientific, as one of the criteria for a scientific method is that the study must be reproducible. This critique is based on an objectivistic scientific paradigm where science consists of statements that can be verified. In contrast to this is the interpretative paradigm where science is understood as the object of continual clarification and discussion (Kvale, 1997). The persona method is as such qualitative; deep knowledge of user needs, attitudes, and behaviour is gathered using qualitative methods. Thus this criticism can be disproven as the critics having misunderstood the starting point of the method.

The method has additionally been criticized for not being able to describe actual people as it only depicts characteristics.

When it comes to implementation, the method is criticized for preventing designers meeting actual users, as actual stories and encounters with real users are assumed to give a better understanding of the users' needs. Yet another objection is that the method does not take into consideration internal politics, and that this can lead to limited use. Lately, the latter has been refuted, as can be seen in the suggested 10 steps to personas that involve the organisation in as many steps as possible.

## 30.3 THE USE OF PERSONAS

I now provide a brief introduction to why the persona method is a useful design tool, what personas are used for, how they are constructed, how to use them, and what to consider in the communication of personas.

In the design process, we begin to imagine how the product is to work and look before any sketch is made or any features described. If the design team members have a number of persona descriptions in front of them while designing, the personas will help them maintain the perspective of the users. The moment the designers begin to imagine how a possible product is to be used by a persona, ideas will emerge. Thus, I maintain that the actual purpose of the method is not the persona descriptions, but the ability to imagine the product. In the following, I designate these product ideas as *scenarios*. It is in scenarios that you can imagine how the product is going to work and be used, in what context it will be used, and the specific construction of the product. And it is during the work with developing scenarios that the product ideas emerge and are described. The persona descriptions are thus a means to develop specific and precise descriptions of products.

A scenario for Virk.dk could be described like this:

Dorte sits at the computer ready to handle the reporting. She goes to Virk.dk and looks at the front page. Dorte tries to take it all in. It is new and thus a little daunting, but at least it looks nice with those colours, she thinks, while at the same time she says to herself: "What do I do now I have installed my signature. I need to find the report form, but where? Well... look, in the middle of the page is a search box, maybe I can try searching for it." Dorte writes "Apprentice refund" in the search field and launches the search. A new screen appears with a short list of hits. At the top is a link to the form she wanted.

Dorte sits at the computer ready to handle the reporting. She goes to Virk.dk and looks at the front page. Dorte tries to take it all in. It is new and thus a little daunting, but at least it looks nice with those colours, she thinks, while at the same time she says to herself: "What do I do now I have installed my signature. I need to find the report form, but where? Well... look, in the middle of the page is a search box, maybe I can try searching for it." Dorte writes "Apprentice refund" in the search field and launches the search. A new screen appears with a short list of hits. At the top is a link to the form she wanted.

FIGURE 30.2: Excerpt of scenario for Virk.dk.

In this case, the designers got the idea that there should be a search field in the middle of the homepage. This would support a shift from information push in the old version of the webpage to form retrieval in the new version.

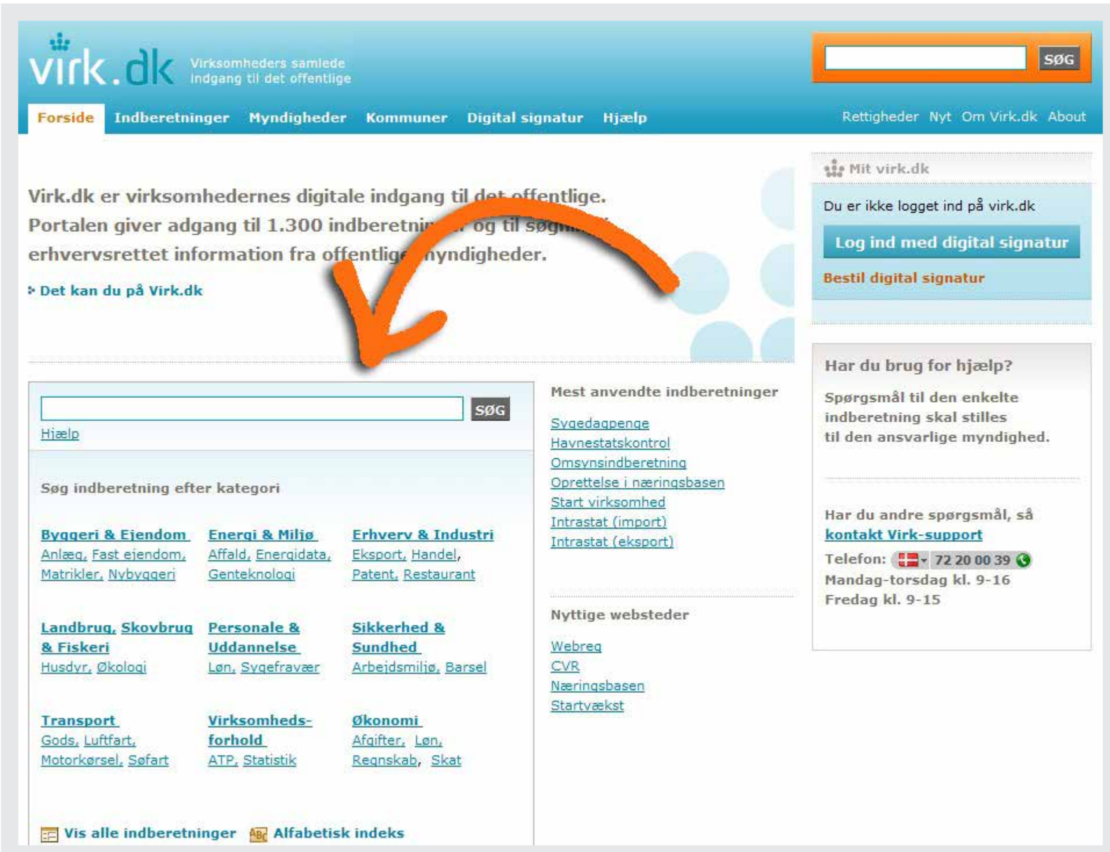



FIGURE 30.3: The front page has an additional search field for Dorte, and focuses on reporting (2008). The latest version of the virk.dk portal has yet again an unclear focus on the front page as it includes both reporting and different guidelines, for example, for company start-ups.



## 30.3.1 Personas for IT and for products

When personas are used for IT system development, it is mainly to explore interaction and navigation. On the other hand, they are not suited to describing what kind of information the system is to contain. For Virk.dk, four personas were created. One of these personas, Jesper, is an accountant. Jesper represents all the users reporting for other companies, not, however, including for example agricultural consultants. When Jesper is used in the scenario, he illustrates what demands those who report on behalf of companies have of the product, both how they think and how they would like to see the information presented. But Jesper does not represent what information and reporting forms should be available on the site. If he did, there would be no reporting forms for agricultural consultants.

In IT systems development, there are specific methods to describe the system navigation and interaction e.g. Unified Modelling Language or the user stories within agile development methods, where the scenarios are described in illustrated and narrative stories. It is quite easy to insert the personas into these methods.

As the persona method developed from a method for IT system development to also working within product development, a shift in the method occurred that has had an impact on how the method is used and what it can be used for. When personas are used for product development, the scenario takes a different path. Here, the users' processes and interactions with the product can more easily be understood as role-playing and story-boarding, and the ways of capturing ideas are less formalized.

### 30.3.2 *4 a*reas of importance, *10* steps to follow

The persona is ideally shaped according to user studies, and used for ideation. But the process varies; sometimes the idea comes first, and the persona is created later and used to verify or enrich the idea (Chang et al., 2008). Studies have shown that one of the main difficulties in the method is to get project participants to use it (Browne, 2011). In the following, I present a process model that sets out to cover this problem throughout the process. It contains four different main parts:

data collection and analysis of data (steps 1, 2), persona descriptions (steps 4, 5), scenarios for problem analysis and idea development (steps 6, 9), and acceptance from the organization and involvement of the design team (steps 3, 7, 8, 10).

The 10 steps cover the entire process from the preliminary data collection, through active use, to continued development of personas. It is an ideal process, and sometimes it is not possible to include all steps in the project (Nielsen, 2011; Nielsen 2012).

*1. Collection of data*. In the first step, you collect as much knowledge about the users as possible. Data can come from many different sources, even from pre-existing knowledge in the organization.

*2. You form a hypothesis*. Based on the first data collection, you form a general idea of the various users within the focus area of the project, including in what ways the users differ from one another.

*3. Everyone accepts the hypothesis*. In this step, the goal is to support or reject the first hypothesis about the differences between the users. This happens by confronting project participants with the hypothesis and comparing it to existing knowledge.

*4. A number is established*. In this step, you decide the final number of personas.

*5. You describe the personas*. The purpose of working with personas is to be able to develop solutions based on the needs of the users, which you do through preparing persona descriptions that express enough understanding and empathy for the readers to understand the users.

*6. You prepare situations*. As already mentioned, the method is directed at creating scenarios that describe solutions. To that purpose, a number of specific situations that could trigger use of the product are described. In other words, situations are the basis, or the precursors, of a scenario.

*7. Acceptance is obtained from the organization.* It is a common thread throughout all 10 steps that the goal of the method is to involve the project participants. This means that as many as possible should participate in the development of the personas and that it is important to obtain the acceptance of the participants of the various steps. That is why all should contribute to and accept the situations. In order to achieve this, you can choose between two strategies: You can ask the participants for their opinion, or you can let them participate actively in the process.

*8. You disseminate knowledge.* In order for the method to be used by the project team, the persona descriptions should be disseminated to all. It is, therefore, important early on to decide how this knowledge is to be disseminated to those who have not participated directly in the process, to future new employees, and to possible external partners. The dissemination of knowledge also includes how the project participants will be given access to the underlying data.

*9. Everyone prepares scenarios.* As previously mentioned, personas have no value in themselves. Not until the moment where the persona is part of a scenario - the story about how the persona uses a future product - does it have real value.

*10. On-going adjustments are made.* The last step is the future life of the persona descriptions. The descriptions should be revised regularly, approximately once a year. There can be new information, or the world could change and new aspects may affect the descriptions. Decisions have to be made whether to rewrite the descriptions, or add new personas, or whether some of them possibly should be eliminated.

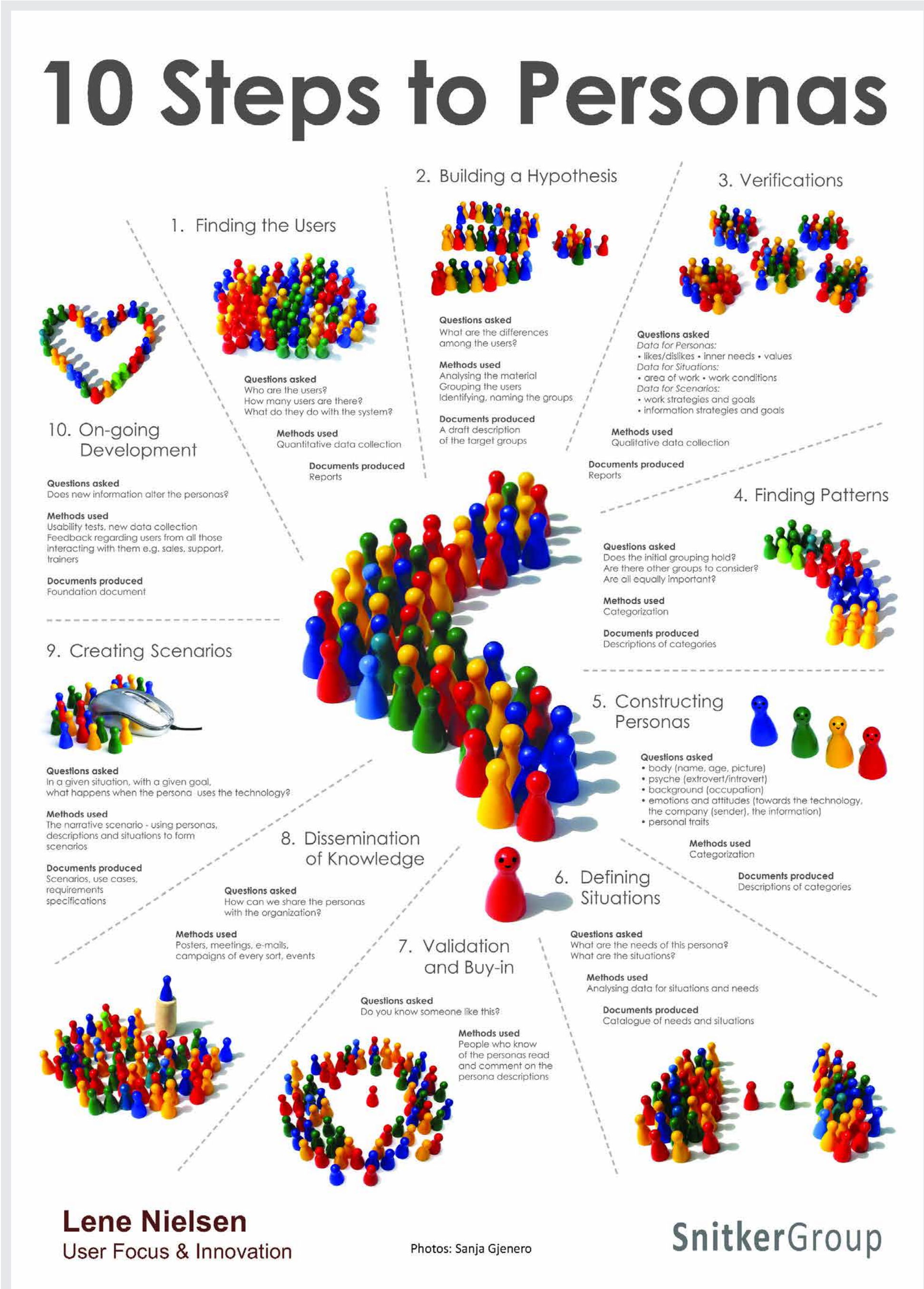



FIGURE 30.4: The poster covers 10 Steps to Personas.

### 30.3.3 Why stories matter

One of the perceived benefits of personas is that they give the design team a mental model of a particular kind of user, which allows the team to predict user behaviour. The personas evoke empathy with users and prevent designers from projecting their own needs and desires onto the project (Floyd et al., 2008). For the persona to evoke empathy, the description needs to be crafted in such a way that the reader can imagine a real person, understand this person's needs and desires, and predict the person's future actions. In the following, I go into more details about how personas and scenarios are tightly interlinked with storytelling to evoke empathy and identification. Here, I am thinking both of the relationship between the fictitious characters and the story and the general narrative structures.

Using narratives is nothing new within IT system development (Madsen & Nielsen, 2006), where stories have been suggested as a starting point to collect data and as a method to theorize over various project types and project phases. At the same time, focus on stories can play a part in providing insight into what goes on outside and below the official course of events. Thereby, the many, often contradictory and competing, stories and interpretations that circulate in an organization can be revealed. Stories can also be used when you want to theorize about organizations, IT systems, and IT system development. An organization can be seen as a collective narrative system where members construct and play out sequences of events on an on-going basis, both individually and together, to be able to remember and to create meaning in past, present, and future events (Boje, 1991; Boje, 1995). These sequences can be used in the process of structuring both the IT system development process and the IT system itself. Here, the narrative contributes to establishing a partnership and a common understanding of the players involved and their goals. This applies in relation to both the development and the presentation of the system to the user (Gazan, 2005). Stories work on several levels, and also provide templates for gathering and analysing empirical data. This happens in interview situations when you need to determine system requirements. The developers can focus on the users' stories about existing and future practice, analyse them, and thus become more aware of requirements (Alvarez & Urla, 2002). These requirements can later on be described in narrative scenarios that are easy to relate to and easy to remember. The scenarios draw on our ability to create meaning individually and together, and to arrange and concentrate information in a narrative form (Carroll, 2000). Subsequently, the stories can be used to analyse the process, the mistakes that occurred, and the political implications of the development and implementation process (Brown, 1998; Brown & Jones, 1998).

## 30.4 FIELD DATA AND THE ENGAGING PERSONA

To describe users as personas helps designers to engage in the users during the entire design process. It enables the design team to engage *in* the user and to focus the design *on* the user. But the descriptions can conflict with the preconceived perceptions the designers have — their stereotypical images.

When we encounter a stranger, we have a tendency to see the person as a stereotype. We do not see the person as possessing a unique constellation of characteristics, but add him to a previously formed category (Macrae & Bodenhausen, 2001). The stereotype is built on knowledge of previous meetings with others and ordered into categories that form the basis for the stereotype. One definition of stereotypes is that they are "socially constructed representations of categories of people" (Hinton, 2000). The stereotypes function as mental pictures for the designers, but being stereotypes, they prevent identification with the described person. The descriptions thereby influence the value of the scenarios as means to investigate and describe a possible future solution.

To move designers away from stereotyping and to get them to engage in the personas puts demands on the descriptions. The field data for personas has traditionally focused on behaviours and demographics (Goodwin, 2008) or goals and tasks in work-related environments (Carroll, 2000). Cooper et al. (2007) list three types of user goals that should be included in the persona descriptions: experience goals, end goals, and life goals. To collect field data for the engaging personas demands an awareness of other kinds of information such as background, psychology, emotions, and character traits.

Within the field of HCI, ethnographic material is collected, interpreted, and communicated, but the act of communicating the data to a design team is often overlooked. The foundation document (Pruitt & Adlin, 2006) is an example of a recommendation for what should be included in a material that can support the personas, but the choice of material does not reflect the process of perceiving the material.

The creative process of writing engaging personas is grounded both in the writer's previous experiences and in the field data. To introduce field data to designers is an act of communication that involves both a selection of the data to present and selecting the form in which it is to be presented. During the presentation, the presenter must be aware of how the data is received and interpreted. As human beings we have a tendency to categorize persons, but by being aware of this and adding information that works against the tendency, we might prevent designers from conceiving stereotypes.

### 30.4.1 Engagement

> INT. - THELMA'S KITCHEN - MORNING
>
> THELMA is a housewife. It's morning and she is slamming coffee cups from the breakfast table into the kitchen sink, which is full of dirty breakfast dishes and some stuff left from last night's dinner which had to "soak".
>
> She is still in her nightgown. The TV is ON in the b.g. From the kitchen, we can see an incomplete wallpapering project going on in the dining room, an obvious "do-it-yourself" attempt by Thelma.

FIGURE 30.5: Beginning of the film script: Thelma and Louise (Khouri, 1990).

But what is it to engage a designer? And how can a persona description evoke empathy? In the following, I present a framework that emanates from a theoretical understanding of fictional "engaging characters" and character building and writing. Furthermore, it encompasses what elements a character description should include and how in the reading process we engage with a character when, as readers, we comprehend a written description of a human as complete. Take a look at the excerpt from the beginning of the film script for *Thelma and Louise*. The description of Thelma is quite vivid and in few lines tells the reader about her state of mind (unhappy), her status (housewife) and her character trait (sloppy).

The character perspective in fiction has commonalities with the persona description as the narrator is withdrawn and invisible. The invisible narrator is also seen in the script example where the description of Thelma is not seen from any specific perspective, but a neutral author comments on the scene by describing details that say something about Thelma and her surroundings.

The persona in the scenario has a function similar to the function of the character in the film script; it is through the description of the character that we understand the character's actions and motivation for action, and it is the character's actions that move the story forward.

### 30.4.2 Levels of engagement

To create an engaging persona is to provide the reader with a vivid description of a user, so vivid that the reader can *identify* with the user throughout the design process. The term "identification" covers the processes of recognition, alignment, and allegiance (Smith, 1995).

- Recognition is the information that enables the reader to construct the character as an individual and human agent.
- Alignment is the process whereby the reader is placed in relation to the character's actions, knowledge, and emotions.
- Allegiance is not only the moral evaluation the reader produces of the character, but also the moral evaluation the text allows the reader to produce. Allegiance depends on access to the character's state of mind and the reader's ability to understand the context wherein the action takes place.

The engagement is first and foremost provided through the ability to recognize — a recognition that originates from both the material presented and the reader's knowledge that goes beyond this. The more material presented the less the reader has to draw on his own experiences.

What is not presented to the reader in the description of the persona, the reader must create himself from inferences, which are added to the story, so to speak.

To enable engagement in the character, the description of emotions as well as of alignment and allegiance is derived from the material. From the reader's point of view, the description of the persona has certain demands, but as the engaging persona is to be part of a scenario, the narration generates demands of the character description too.

In order for the designer as reader to engage in the persona, the text must give access to:

- Information that enables persona construction;
- Information about the emotional status of the persona;
- Information about context that enables an understanding of the character.

### 30.4.3 Flat character or rounded character

To the reader or viewer of fiction, the two basic types of stories are well known. The plot-driven story is seen in most action films where the hero has very few character-traits and the story moves from one action to the next. The character is defined as a flat character. In contrast to this, the character-driven story has a character that is described with several traits, and it is the development of the character that moves the story forward. This character is defined as a rounded character.

The difference between the characters in the character-driven and the plot-driven stories are the number of traits. In the plot-driven story, the character has a limited number of traits, and they function as catalyst for actions. This makes the character highly predictable and creates a flat character in a plot-driven story (Chatman, 1990).

In the character driven story, the character has multiple traits. The tension between the character traits creates unpredictable actions and this determines the story development.

### 30.4.4 Construction of the rounded character

A description of a human being should include physiognomic dimensions as well as sociological and psychological dimensions. Each of these has an impact on the character's behaviours and actions. The descriptions of physiognomy, sociology, and psychology enable an understanding of the motivation behind the persona's actions in the scenarios. The dimensions encompass both past and present, both the self and relations to others.

"If we understand that these three dimensions can provide the reason for every phase of human conduct, it will be easy for us to write about any character and trace his motivation to its source." (Egri, 1960: p. 35).

Thus the character possesses both personal (inner) and inter-personal (social, public, and professional) elements. It is the diversity in experiences that constitute the character. All characters have personal needs and goals as well as interpersonal wishes and professional ambitions. These help define the character and create its own demands, restrictions, and privileges.

This, as well as the understanding of the rounded character, has several implications for the creation of the persona description:

- The character develops in the story.
- The character has two or more traits that interact.
- The character belongs to a specific time and specific culture and interacts with the culture.

The persona description is characterised by:

- More than one character trait;
- Psychology, physiognomy, and a social background;
- Personal needs and goals, interpersonal wishes, professional ambitions;
- Context.

Furthermore the persona belongs to a specific time and specific culture and interacts with the culture, and the persona develops in the scenario.

The understanding of the rounded and engaging characters puts demands on the data from the field studies. Observations of work processes and segmentation of users as skilled versus non-skilled users do not provide the writer with the necessary information to describe the character traits and psychology of the persona. As the rounded character incorporates personality, emotions, and actions, these elements should also stem from the field data. During research, information about the users' surroundings, character traits, emotional status, and background, as well as common demographics, actions, goals, and tasks should be observed.

## 30.5 SCENARIOS AND STORIES

The move from creating engaging personas to scenarios reflects the move from a static character description to an active character in a story — a narrative. In the following, I draw on an understanding of narratives as both a process (mental story construction) and a product, both performance and text (Ryan, 2003; Boje, 1991).

There is an on-going discussion of what a narrative is, but it is agreed that it includes time — it is a way for humans to organize time. Is a narrative the barest organization of time as suggested by Abbott (2002) e.g. "She took her bicycle"? Or does it include causality, as suggested by Cobley (2001) e.g. "She took her bike and was hit by a bus"? I will not participate in the discussion, but take a pragmatic standpoint from a scenario point of view. In the scenarios, the narrative must include causality, as the focus is on the relationship between how the user's action gives rise to system reaction and how system action provokes user reaction.

There is a distinction between narrative and story. The narrative is the organization of events, while the story includes prior events or events the reader must assume or guess.

**Narrative:** A movement from a start point to an end point, with digression that involves the showing or telling of story events. Narrative is a *re*-presentation of events and, chiefly, *re*-presents space and time.

**Story:** All the events which are to be depicted in a narrative and which are connected by the means of a plot (Cobley, 2001).

For a text (in the broadest sense of the word) to qualify as a narrative, it must (Ryan, 2003):

- create a world and populate it with characters and objects; the world must undergo changes of state that are caused by non-routine physical events: either accidents/happenings or deliberate human action;
- allow the reconstruction of an interpretive network of goals, plans, causal relations, and psychological motivations around the narrated events.

### 30.5.1 Stories and meaning

When we experience incidents, we try to understand and extract a meaning from the incidents. This meaning can either have the construction of a story with sequences organised in a narrative and linear structure or as meaning organised in a logical structure (Bruner 1990). When we read a scenario, a similar process begins. We are handed story elements that make us put together a narrative. Some information we do not receive as story elements; they are to be inferred from our expectations, knowledge of the area, and cultural background.

The creation of the narrative is an inter-subjective process constructed by the reader from presumptions and inference. An example: The woman takes the knife. The man hits the woman.

We infer a causal connection between the two elements of action, and we assume that the two actions will spur further action.

The narrative becomes different if the story-elements are the same, but are presented in reverse order, as in this example: The man hits the woman. The woman takes the knife.

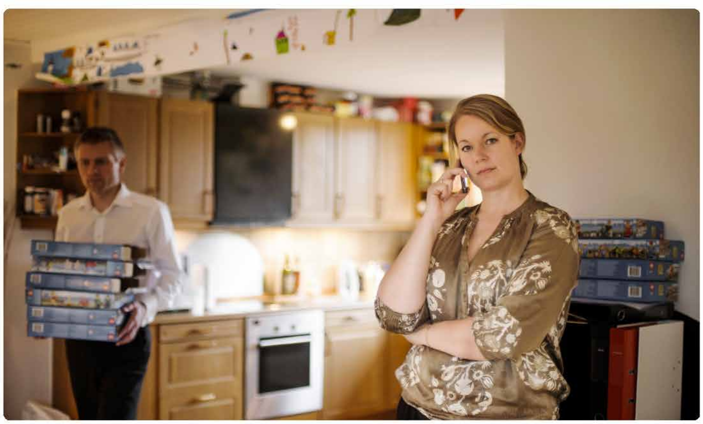
### 30.5.2 Personas in stories - scenarios

In IT system development, the persona description is used as the foundation for outlining a scenario that investigates the use of an IT system from the particular user's point of view. The persona method authors suggest different types of persona-scenarios. Cooper et al. (2007) suggest a progression from initial, high-level scenarios to more and more detailed ones with increasing emphasis on the user-product interaction. In order to describe this progression, they distinguish between *problem scenarios*, which are stories about a problem that exists prior to the introduction of technology, and *design scenarios* that convey a new vision of the situation after the introduction of technology. Pruitt & Adlin (2006) refer to Whitney Quesenbery's (Quesenbery 2006) definition of different types of personas and to scenarios with different levels of detail placed in a continuum between evocative and prescriptive scenarios as well as along the development process. Mulder & Yaar (2006) focus exclusively on web development and propose one type of scenario that describes a persona's journey through a website. The authors provide different lists of elements that could/should be included in a 'complete' or 'good' scenario. Below is presented an overview of these authors, the scenario elements they propose, and the scenario element definitions they give, if definitions are given (Madsen & Nielsen, 2009)

| | Cooper (Cooper et al., 2007) | Quesenbery (Quesenbery, 2006) | Pruitt & Adlin (Pruitt & Adlin, 2006) | Mulder & Yaar (Mulder & Yaar, 2006) |
|---|---|---|---|---|
| **PERSONA** | A persona encapsulates a distinct set of behaviour patterns | *Characters* The persona is the main character, but other characters might also be involved. | Specified *user* A persona. | Document the mission from the persona's point of view. Others can be around, influencing the decisions. |
| **CONTEXT** | In what *setting* will the IT system be used? | | | *Set the scene:* Where is the persona? When does it happen? Who else is around? |
| **BEGINNING** | | *Context/situation* The beginning of the story, motivation for what happens, and a focus on what the persona is trying to do. | Particular *task* or situation. | *Establish the goal or conflict.* Conflict: inner or outer conflict triggers the visit to a website. |
| **ACTION** | Is the persona *interrupted* frequently?<br>Will the IT system be used for extended amounts of *time*?<br>Will other IT *systems* or products be used?<br>What are the *primary activities* the persona needs to perform to meet his or her *goals*?<br>How much *complexity* is permissible, given persona skill and frequency of use | *Plot/action* Focus on what happens during the scenario and what influences the decisions the persona has to make.<br>Facts Gathered from user research. | *Features* or *functionality* References to specific features or functionality the persona will need and/or use.<br>Procedure or *task flow information* | *Overcome crisis* along the way. Describe intermediate steps and decisions.<br>Does the persona use other sites, email, phones etc.?<br>How does the persona feel about the experience? |
| **ENDING** | What is the *expected* end result of using the IT system? | *Resolution* The ending situation and a focus on what has changed during the story. | Clearly defined *outcome* or *goal* for the task. | *Achieve resolution*<br>Reach denouement - what happens after the resolution? |

TABLE 30.1: An overview of the various scenario elements according to different authors.

The table shows that the lists of scenario elements are somewhat similar, but also that only Quesenbery (2006) and Mulder & Yaar (2006) explain the elements that should be included in a scenario — and this only in a brief manner. Mulder & Yaar (2006) state that the scenario elements they outline are the classic components of storytelling. However, they do not explain what classic storytelling is. In general, the persona literature is clearly inspired by, but does not explicitly refer to, narrative theory as an established knowledge base and source of already defined (if controversially discussed) key concepts, such as story elements. It is relevant to look more closely at the narrative aspect of persona-scenarios and to draw more explicitly on narrative theory in doing so.



"If you don't have to enter anything and data comes from a register, then it's even smarter - you don't have to do anything. But it's important that you get a clear message from the system and know your status – or that you get a receipt"

Motivation for use

Camilla and Jesper would like to be able to use online self-service solutions outside ordinary opening hours. They want to control the process themselves and they get frustrated if they cannot understand their rights and the process of the service. (11)

Accessible for all

Camilla and Jesper could have a handicap. The solution should be able to cater for special needs and for users that use compensating tools, this can be ensured by applying the guide for web accessibility (WCAG).[12]

Examples of self-service situations

- Sign up for daycare
- Change kindergarten
- Sign up for school and after-school
- Submit tax
- Report mileage to the Danish tax authorities
- Check estate information
- Change of address when moving
- Order new license plates
- Report for unemployment compensation
- Repay Student Loans
- Sell a car and buy a car

Notes

1. The persona descriptions are based on data from Statistics Denmark, Index Denmark, Gallup Marketing and a special query from Geomatics Conzoom 2011. This query concerns: use of IT, education, income, age, location, attitudes regarding security on IT, and attitudes towards societal development. Data concerning relations to the public sphere and self-service is collected from qualitative interviews.
2. All quotes originate from qualitative interviews (SnitkerGroup, 2012)..
3. For the group "The Autonomous" the household income is above 770.000 d.kr, 38% have a long or a medium long education and 29% have a vocational education.
4. Within the age group 25-44 90% holds NemID (Statistics Denmark, 2011)..
5. For the group "The Autonomous" 53% uses E-boks (Conzoom, 2011). 90% of those with a longer education use an Internet bank (Statistics Denmark, 2011).
6. In the municipality of Lyngby-Taarbæk 45,1% of the families with children age 0-6 years has signed up for Digital Mail (Danish Agency for Digitisation, 2012).
7. 77% of "The Autonomous" shop on the Internet. 70% feel safe shopping online (Conzoom, 2011).
8. 84% of the Danes between 25-34 years and 71% of the Danes between 35-44 years use online social networks. For both age groups 28% use professional online social networks. For those with a longer education the numbers are: 62% use online social networks and 31% use professional online social networks (News from Statistics Denmark, 403, 2. Sept., 2011).
9. 55% of "The Autonomous" have 1-2 mobile phones in the household, 42% have 3 or more (Conzoom, 2011).
10. 61% of employed persons with income above average and 58% of self-employed have used mobile Internet outside the home within the last 3 month .
11. 90% of Danes between 25-44 years have tried to search for information on a public homepage. 76% of Danes between 25-34 years and 78% of those between 35-44 years have tried to upload information via a public homepage (Statistics Denmark, 2011).
12. www.w3.org/Translations/WCAG20-da/

borger.dk

FIGURE 30.6: Personas for borger.dk. Each persona description has examples of situations that act as a starting point for scenarios. In this case, e.g. change of kindergarten, and change of address. Note that in this case, the persona is a couple.

### 30.5.3 The story structure

Basically, a story includes setting, goal, plot, and solution. It evolves in a dramaturgy of beginning, middle, and end. For the scenario, the dramatic elements are:

- The beginning presents the user and what the persona wants to achieve.
- The middle describes what the user does, e.g. the navigation and the information that is offered. And it describes the persona's motivation for pursuing the goal.
- The end describes whether the persona succeeds in his or her intentions.

The creative part of scenario construction is crucial; it is during this process that the field data evokes design ideas. Whether writing or drawing, producing a scenario can take place in a shared setting and as an act of communication. The scenario explores the creative situation, and this can be compared to the productive notion Tolstoy had when writing Anna Karenina:

"(...) after Vronsky and Anna had finally made love and Vronsky had returned to his lodging, he, Tolstoy, discovered to his amazement that Vronsky was prepared to commit suicide." (Abbott, 2002: p. 18)

It is the acquaintance that Tolstoy has with the Vronsky character and the narrative that in itself provides the action. In this instance, it reflects the circumstances and the traits of the character, but when the story acts as a motor for creativity, it can both provide descriptions and solutions and, as here, also carry the writer away.

As the creative process has this ability, it becomes necessary for the producer of the scenario to look for consistency and closure in both the description of the persona and the scenario. In other words, the producer needs to look upon the scenario from a reader's point of view to try to understand the story as an outside reader might understand it.

### 30.5.4 Story elements: Events

Events are essential to stories. A narrative is made up of both constituent events that are necessary for the story and supplementary events that do not seem essential, but add flavour and enrich the story (Abbott, 2002).

Setting is another common ingredient in stories. The setting is optional; "she took her bike and was run over by a bus" does not include a setting, but to a scenario, the setting is inevitable. It is the setting that pinpoints where the use takes place, the surroundings that may influence the use, the time of day, and other elements of context that might influence the use.

> Dorte sits at the computer ready to handle the reporting. She goes to Virk.dk and looks at the front page. Dorte tries to take it all in. It is new and thus a little daunting, but at least it looks nice with those colours, she thinks, while at the same time saying to herself: "What do I do now," I have installed my signature. I need to find the reporting, but where?" Well... look, here in the middle of the page is a search box, maybe I can try searching for it" Dorte writes "Apprentice refund" in the search field and launches the search. A new screen appears with a short list of hits. At the top is a link to the form she wanted.

Figure 30.7: Excerpt of scenario for Virk.dk, 2007.

If we again look at the scenario for Virk.dk, the setting is known from the persona description. The scenario includes the following events:

- Dorte writes "virk.dk" and presses the enter key.
- The webpage loads.
- She searches for a form.
- She gets a list of hits.

These are the events that lay the foundation for the system design.

### 30.5.5 Story elements: Closure and resolution

When a narrative resolves a conflict, it achieves closure (Abbott, 2002). Closure can be both for a single event and for the whole story. Resolution is one way of obtaining closure. When we read a story, we want to have closure to get answers to questions and to experience the end. We have a desire for the end. As Brooks puts it: "...narrative desire is ultimately, inexorably, desire *for* the end" (Brooks, 1984: p. 52). As we write the scenario, we do not know the end, but we have a desire to find the end through the creative process. During discussions and writing, endings will be created; if open-ended, the scenario can be an agenda for dialogue.

Scenario reading differs from scenario writing. As a natural cause of the reading process, closure and resolution is craved. If it is not provided, the reader will make it up (Abbott, 2002) — in the scenario example, the reader depicts Dorte in a setting: is she in her office — or sitting with her laptop in front of the television? If the scenario does not provide the information, the reader will make the setting up in the urge to obtain closure.

### 30.5.6 Story elements: Voice

The voice — whom do we hear — can be either a first person narrator or a third person narrator. The first person narrator "I took a bicycle" is limited to the view of the person, while the third person narrator "she took a bicycle" has the possibility to include comments that the person cannot know of — "she took a bicycle, but what she didn't know was that the evil person had put a GPS on it to track her traces".

Some suggest use of a first person narrator, e.g. (Mikkelson & Lee, 2000). Even though this provides information about the language used by the persona, the downsides are bigger. The possibility of adding omnipotent comments to the scenario should not be ruled out as they can function as containers of information that are not accessible for the persona. If dealing with scenarios for more personas, it can be very difficult to distinguish one character from the other with the use of "I". Furthermore, it takes a very skilled writer to imitate and create a consistent language. This can be experienced in the scenario excerpt where the dialogue is very extensive and not as subtle as real-world dialogue.

### 30.5.7 Story elements: Plot

Plot is inevitable in a story; it is the linking of events in the plot that keeps the story moving (Cobley, 2001). Both plot-driven and character-driven stories include action; what separates them is the depth of the persona portrait. In the character-driven story, the character is seen as a personage rather than somebody who is the product of the plot and just participating in the story development. Instead, it is the character development that creates the story development, and the story development that spins the plot. If the designer is to engage with the persona, the scenario should include a strong central character with goals and desires that need fulfilment during the story, thus resembling the character-driven story.

### 30.5.8 Story elements: Obstacles

Understanding the obstacles of the users when using the product is essential in order to develop and redesign products. Often, the obstacles become visible during the collection of data, and they can stem from the system itself and from the surroundings. If we look again at the excerpt of the scenario for Dorte, the obstacle of installing a digital signature is not examined. It is a pre-requisite that she has installed the signature, and the story begins from this point. But how likely is it that she can do this herself when she needs help from her son to do her first digital reporting?

You often see scenarios, especially in IT, where just giving the persona an IT system solves all problems. I call this type of scenario “happy scenarios” as they always have a happy ending. They lack obstacles, and the solution is always that the IT product will make everything much easier. When you read this type of scenario, they seem unrealistic because of the lack of obstacles.

### 30.5.9 An overview of the scenario elements

The overview below is inspired by Jean Mandler’s reflections on the story form (Mandler, 1983). The story begins with a setting in which characters, location, problems, and time are presented. After this presentation, one or more episodes follow, each having a beginning and a development towards a goal. In the opening episode, the character reacts to the introductory events, sets a goal, and outlines a path to reach the goal. Each episode focuses on the goal, the attempts to reach the goal, and the obstacles in the way of reaching the goal. The attempts are understood as the causes of the outcome. Each episode links to the overall story, thereby building up the plot.

| NARRATIVE ELEMENTS | NARRATIVE ELEMENTS IN A SCENARIO |
|---|---|
| **Character(s)**: a protagonist as well as minor characters. A character can be any entity that has agency, that is, involved in the action. | In scenarios, the persona is the protagonist. (In scenario-based design, the main character and protagonist is the IT system.) |

| | |
|---|---|
| **Time**: both the time in which the actions take place, e.g. the future, and the story development over time - beginning, middle, and end. | Most scenarios are set in present time, but they can also concern a distant future. The story time can last minutes, days, months, etc. |
| **Problem**: a loss, a need, a lack of something, an obstacle to overcome, a conflict. | The persona has a problem. |
| **Setting**: presentation of characters, location, problems, and time. | The scenario begins with a presentation of the persona, his or her problems, the place where the action takes place and the time (present time/distant future). |
| **Opening episode**: the character reacts to the problem, sets a goal, and outlines a path to the goal. | The persona defines the goal and starts to act. |
| **Episodes**: development toward the goal. Episodes consist of:<br>• Beginning<br>• Attempts<br>• Events (accidents, obstacles, happenings, deliberate human actions)<br>• Development | The scenario develops through a sequence of episodes that concern the problem, the goal and the attempts to reach the goal, the events involved in these attempts, and the obstacles hindering fulfilment of the goal. |
| **Resolution**: the problem is solved and the goal is reached - or it is not. | There are two types of scenarios - one where the problem is solved and the goal is reached, and one where they are not. |

| | |
|---|---|
| **Plot**: the linkage and order of the episodes. | Most scenarios are presented in a linear manner, without deviations from the story line. |
| **Overall story**: starts with a beginning, goes through a middle, and arrives at the end. The overall story is sensitive towards what is considered ordinary social practice within a given culture and explains deviations from accepted social practice. | Each episode links to and has to be meaningful in relation to the overall story.<br>The scenario has to explain why non-routine actions and events happen and how they are dealt with. |
| **Narrator's perspective**: The narrative is told by someone. | Most scenarios are told in third person allowing the narrator to be omnipotent. |

TABLE 30.2: An overview of the story form and a 'translation' hereof to a scenario context (Madsen & Nielsen, 2010).

### 30.5.10 Benefits and pitfalls

To sum up, the benefit of the scenarios are that they are specific and use a language that is easily understood and accessible for both users and designers. This is a contrast to other models that require knowledge and expertise in order to be understood. The scenario enables a design process focused on *use* and explains vividly why a *system is necessary*. The scenario enables an understanding of the experiences that most likely results in a successful accomplishment of the *user's goals* and it offers a *task-oriented* decomposition (Sutcliffe, 2003; Bliss, 2000; Jarke, 1999; Kyng, 1992).

This overview suggests a shared understanding of the scenario method as a means to provoke discussions and generate ideas in a specific language easily shared by users and designers. It enables experiences to be shared and anchors the empirical data in a specific form.

The downsides of the scenario are that it can confirm beliefs and create obsession with details, and it can create a false reassurance, as it is hard to know when the design area is covered. The downside is also a lack of clarity in evidence of field data and a lack in focus on the user.

The benefits of the scenario are:

- The scenario enables specific knowledge and supports reflection in action.
- It supports communication and shared understanding between the members of the design team and between users and designers. It can help design knowledge to accumulate across problem instances when categorized and abstracted.
- It supports idea generation, including abandoning of design ideas and is easily revised and written for many purposes and at many levels.
- It constitutes a theoretical anchoring of an empirical "chaos".

(Bødker, 1999; Bødker & Christiansen, 1997; Carroll, 1999, Carroll 2000; Erickson, 1995; Sutcliffe, 2003).

The downsides of the scenario are:

- The scenario can create a false sense of assurance that all aspects are covered by a small number of scenarios and can supply minimal evidence to confirm a belief.
- As the scenario is detailed, it can bias people away from the big picture and create obsession with unnecessary details.
- The scenario method lacks clarity in defining the user: it does not capture the essence of the user as representation is covered by

attributes such as age, job, and title and might not faithfully represent the user's tasks and contexts or the user's interests.

- The method lacks focus on the user and does not create engagement.
- The scenario design tradition lacks evidence of how data is gathered and on which basis the scenarios are formed.

(Grudin & Pruitt, 2002; Mikkelson & Lee, 2000; Nardi, 1996; Nielsen, 2003; Sutcliffe, 2003).

Alan Cooper (1999) links personas and scenarios. He describes the scenario as the investigation of tasks. "A scenario is a concise description of a user using a software-based product to achieve a goal" (Ibid p. 179), where the goals stem from the persona description.

Even though it seems natural that there should be a link between personas and scenarios, they have often been viewed as separate methods.

## 30.6 A USEFUL TOOL AROUND THE GLOBE

In 2010, Forrester claimed that a redesign with personas can provide a return of investment on up to four times (Drego & Dorsey, 2010). In a later study, it was demonstrated that a number of areas affect the success of the method (Browne, 2011). The criteria of success are:

Personas should be used for identifying concrete decisions and products that should be based on personas.

- Qualitative research should have high priority.
- The persona descriptions should be properly made.
- The descriptions should be used for design decisions.
- The descriptions should be evaluated and updated at regular intervals.

Reported failures seem to be due to:

- Sloppy research and data gathering.
- Many do not understand the method and think it is the same as demographically based segmentations and customer profiles.
- Too few resources are put into the project.
- The quality of the persona descriptions is inconsistent. Some are well-written stories and engage stakeholders using supportive material. Other descriptions are unrealistic and badly written and do not think of those using the descriptions.
- Those who are to use the descriptions find it hard to understand how to use them.

Similar studies have been carried out in Denmark in 2009 (Vorre Hansen, 2009) and 2011. These studies showed that there is a difference between consultant agencies and in-house development as some agencies have a world-view or a lens with which they collect and analyse the data, e.g. motives and barriers for product use or shared values between sender and recipient. For in-house development, the specific values seem not to be present. One finding was that those interviewed in the study assessed that personas should be checked regularly and updated if necessary every 1-2 years.

The interviewed judged the value of the method as well as the challenges. From the studies, patterns of what the companies use the method for appeared. Personas are used for a variety of purposes and with different values:

- *In concept and product development*, to maintain focus on the users in the entire development process. The personas can be integrated directly in the development processes, in preparing test scenarios and in concept testing. Some companies use them for requirement specification.

- *As a strategic tool*. Some companies use them as a strategic tool to target future user groups. The method makes it possible to make strategic decisions regarding target groups. The companies are able to identify both their primary target group and groups that are either secondary groups or*not*to be addressed.
- They are used for *recruitment* of users for usability tests, interviews, and focus groups, and for preparing test scenarios and questionnaires.
- The method *offers a common communication platform* for the company or internally in the project group. It ensures that discussions are based on a common understanding of the user and not based on pre-existing understanding of and personal experience with the users.
- It *provides a qualified understanding* of the users. The method communicates data and thus increases the internal knowledge about the target groups of the company.
- The method *shifts* focus from the well-known users to the lesser known, thus ensuring that the target groups that the company knows less about are also included in the deliberations of the projects.
- It can *focus and validate* the final product. By including the users early on in a development process, the likelihood that there will be recipients for the product is greater.
- The method creates *documentation and argumentation for specific solutions*. To be able to refer to a specific persona and the underlying data is part of supporting the choice of one solution over another.
- It *supports working across* departments. Especially in larger organizations, the method can contribute to abolishing “silo” thinking.

When focus is shifted from the organizational structure to the users, the method makes different departments collaborate.

- It has a *long shelf life*. Personas can be included in more and in new projects long after their development.
- *Surprising areas of application*. The persona descriptions can be used in areas that are unexpected e.g. to develop test scenarios and questionnaires.
- A number of challenges are reported, especially in the following areas:
- Making the method *visible* in the organization. It can be difficult to disseminate knowledge about and ownership of the method to other departments in the company.
- It is challenging to get *supporting knowledge* about the persona method as there is no exchange of experiences across companies and there is no platform for knowledge sharing.
- It is difficult to *differentiate the communication* and to operationalize the personas to various work groups; for example, some groups might emphasize that they have insight into the data behind the personas, while other groups dislike the fictional elements. There is a lack of tools to develop differentiated forms of communication.
- Often, the method is *dependent on individuals*, and it is a challenge to anchor the method in management. This creates a feeling that the method depends on individual members of staff, and if they disappear from the department, the method will not survive.
- It may also be difficult to *get external suppliers to use* the method and to communicate personas to them. It is important to get the suppliers to use or at least know the method as in certain contexts

it is the subcontractors that develop the final product.

- *Maintaining the persona descriptions* is also a challenge. Often, there is no one employee responsible for updating the personas so the descriptions are updated at random. Funding for this might also be difficult as there is no budget for updating, but only a budget for the actual persona development project.

In the book “Personas - User Focused Design” (Nielsen, 2012), I asked 6 UX professionals from Australia, Brazil, Finland, India, Russia, and UK, to give a short overview of the status of persona use in their country. The common experience is that it is often large companies that are drivers: banks, newspapers, insurance companies, or government bodies. Some companies report from product design (Japan, India), while most use still seems to be within IT-related areas. Few companies develop their own method, and many refer to Alan Cooper and the goal-directed design method. The Russian UIDesign Group had to translate and define the term of personas into Russian when they started and has now developed its own method that has a strong focus on the use of personas for requirement gathering. Around the globe, there seems to be a uniform format of the descriptions, with photos and short descriptions of an individual character. While most companies report that the use of the method is still in its infancy, Daishinsha Inc. in Japan has used the method for product innovation since 2000, following an introduction from Forrester research and Alan Cooper.

## 30.7 FUTURE DIRECTIONS

As we have seen, the persona method is not a uniform method, and it is used in different ways. With a shift from a method used for IT systems design to include more areas such as marketing and communication, the method is constantly evolving. Some companies do not fully understand the potential of the method as a design method and instead devise marketing archetypes and call these personas. I hope that in the future much more focus will be on use, on the scenarios, and on a better understanding of how data collection should be part of proper persona descriptions.

One constant in the persona method has so far been its focus on users and not on including users in the design process, as can be seen in participatory design and co-design (see e.g. Ned Koch's chapter on action research). In the following, I present a novel way of using personas for innovation and ideation where the users are involved in the design process — personas as part of role-playing scenarios with users (co-design). Furthermore, I present an experiment with actors acting as personas to help designers become immersed in and understand the daily lives of the personas (enactment).

### 30.7.1 Personas and co-design

Arla Foods a.m.b.a., the leading dairy company in Scandinavia, wanted to innovate within the, until then, unknown area of canteens. For the purpose of creating new products from user knowledge, an innovation process was launched. It consisted of: scientific data gathering, customer data gathering, and data analysis. From the analysis, two personas were produced. The material was used in an innovation workshop lasting two days. The participants were to use the persona descriptions in various scenarios and come up with product ideas based on the scenarios.

The innovating participants were concept developers, marketing managers, engineers, and canteen managers. Even though the canteen managers came on the second day of the workshop, they entered the groups without hesitation and became engaged in the creative process. It was easy for them to relate to the persona descriptions and they felt on equal footing with the designers. This resulted in more than twenty ideas out of which four were picked for further development.

Another case of users innovating with personas is from a student session. This case had only one participant, but proved as successful as the first industrial case. The aim was to develop a tool that could support communication between soccer trainers, children, and parents. Prior to the session, data was gathered from observations and focus groups.

From this, two personas that had different behaviour and media use were created as well as a number of scenarios that varied in situation and context. The user was asked to go through all the scenarios from the point of view of the two personas, with the intention of creating novel solutions.


Figure 30.8: The user explains to the moderator how the persona will act in the given scenario.



The participant, the mother of a soccer-playing child, had no problem in switching between the two personas, even though only one resembled herself. She was able to draw on her knowledge of other parents and their preferences and behaviour, but when she acted as the persona that resembled herself, she often commented on the likeness, how she herself would react, and her own needs.

The two cases show how users 1) are able to act as personas and be as creative as professional designers 2) use their understanding of the area in focus to create scenarios from both the perspective of personas that are similar to them, and from personas that are different from them because they are familiar with different behaviours within the given design area.

It also shows how the users immediately are able to role-play through the scenarios and to do this both alone and together with designers and other project participants.

The use of personas enables project participants to discuss from the same understanding of context and needs and at the same time allows the users to enter the discussion as experts and relate to the innovation from their concrete knowledge of culture and work tasks. This way, the personas take the focus away from the users as the single domain experts — as is seen in participatory design — and aligns designers and users in the role as innovators.

### 30.7.2 Persona actors

A recent experiment by Line Mulvad, shows how actors can perform as personas and thereby enhance the engagement and understanding of the users. For the website borger.dk that is aimed at all citizens in Denmark, six personas were developed during spring of 2012. The personas vary in their knowledge of what it takes to be a citizen, their understanding of the public sector, their use of and competences within IT, and their use of digital self-services. As an experiment, Line Mulvad, producer and actor, took two persona descriptions as the point of departure - a male carpenter, age 44, single parent who found reading long text difficult; and a female healthcare assistant, age 56 and originally from Bosnia, thus speaking Danish with an accent. From the descriptions, the producer and the two actors improvised a series of scenes that introduced the personas as characters and a couple of scenarios that showed the problems they have in using public websites — for both, when it comes to understanding the language, and for the woman, the understanding of the systems. All was filmed and edited into two small films of app. 4 minutes length. The films had a dogma style with hand-held camera and the actors talking to the camera, which gave an authentic impression.

Figure 30.9: The actor Jeppe Christoffersen plays a carpenter who has problems with reading long texts and therefore wants to be absolutely sure he has done it right when he reports digitally.



The films were shown to an audience, and the following discussion showed that almost all felt that they got a better understanding of the personas from the movies than from a written description; they felt it was easier to engage in the personas and come up with design ideas. Also, the spectators never felt that they were watching actors, but were convinced that they were watching real users. A downside of this method is the second part of the movie, which incorporated a present-day scenario. The scenarios gave the film action and plot, but as the present day scenarios will change when redesign is being performed, they do not have a long-lasting effect. In the future, we will have to test the difference between the written persona descriptions and the enacted descriptions, and work out how to incorporate action and drama without focusing on problems from the present day scenarios. But by using actors instead of real users, it becomes possible to focus on exactly the information and problems you want to put forward.

## 30.8 WHERE TO LEARN MORE

There are very few books written on personas, but many papers based on single case studies. Here is an overview of the books I have come across.

**The goal-directed perspective:**

- *Cooper, 1999: The Inmates Are Running the Asylum* - the first book to mention the concept of personas.
- *Cooper, 2007: About Face 3.0: The Essentials of Interaction Design* - has two chapters on personas and scenarios.
- *Goodwin, 2009: Designing for the Digital Age: How to Create Human-Centered Products and Services.* - gives an overview of goal-directed design including personas.

**The role-based perspective**

- *Pruitt and Adlin, 2006: The Persona Lifecycle: Keeping People in Mind Throughout Product Design* - a thorough introduction to all steps in personas and scenario construction.

**The engaging perspective**

- *Nielsen, 2004: Engaging Personas and Narrative Scenarios* - a PhD dissertation, rather heavy on theory.
- *Nielsen, 2011: Persona: Brugerfokuseret design.* - introduces the

10 steps, step by step. (in Danish).

- *Nielsen, 2012: Personas - User Focused Design*. Human-Computer Interaction. Springer. - introduces the 10 steps, step by step.

**Web-based design:**

- *Mulder and Yaar 2006: The User Is Always Right: A Practical Guide to Creating and Using Personas for the Web*. - focuses solely on web

## 30.9 REFERENCES

*Abbott, H. Porter (2002): The Cambridge Introduction to Narrative. Cambridge University Press*

*Adlin, Tamara and Pruitt, John (2006). Why is it so Hard to Make Products that People Love?. Retrieved 19 December 2012 from AIGA: http://www.aiga.org/why-is-it-so-hard-to-make-products-that-people-love/*

*Alvarez, Rosío and Urla, Jacqueline (2002): Tell Me a Good Story : Using Narrative Analysis to Examine Information Requirements Interviews during an ERP Implementation. In Data Base For Advances In Information Systems, 33 (1) pp. 38-52*

*Blythe, Mark and Wright, Peter C. (2006): Pastiche scenarios: Fiction as a resource for user centred design. In Interacting with Computers, 18 (5) pp. 1139-1164*

*Boje, David M. (1991): The Storytelling Organization: A Study of Story Performance in an Office- Supply Firm. In Administrative Science Quarterly, 36 (1) pp. 106-126*

*Boje, David M. (1995): Stories of the Storytelling Organization: A Postmodern Analysis of Disney as "Tamara-Land. In Academy of Management Journal, 38 (4) pp. 997-1035*

*Brooks, Peter (1984): Reading for the Plot: Design and Intention in Narrative. Harvard University Press*

*Brown, Andrew D. (1998): Narrative, politics and legitimacy in an IT implementation. In Journal of Management Studies, 35 (1) pp. 35-58*

*Brown, A. D. and Jones, M. R. (1998): Doomed to Failure: Narratives of Inevitability and Conspiracy in a Failed IS Project. In Organization Studies, 19 (1) pp. 73-88*

*Browne, Jonathan (2011). Customer Experience Professionals: Break Through Persona Ob-*

*stacles Now. Forrester Research http://www.forrester.com/Customer+Experience+Professionals+Break+Through+Persona+Obstacles+Now/fulltext/-/E-RES58814*

*Bruner, Jerome (1990): Acts of Meaning: Four Lectures on Mind and Culture. Harvard University Press*

*Bødker, Susanne (1999): Scenarios in User-Centered Design - Setting the Stage for Reflection and Action. In: HICSS 1999 1999.*

*Bødker, Susanne and Christiansen, Ellen (1997). Scenarios as springboards in design of CSCW.*

*Carroll, John M. (1999): Five Reasons for Scenario-based Design. In: HICSS 1999 1999.*

*Carroll, John M. (2000): Making Use: Scenario-Based Design of Human-Computer Interactions. MIT Press*

*Chang, Yen-ning, Lim, Youn-kyung and Stolterman, Erik A. (2008): Personas: from theory to practices. In: Proceedings of the Fifth Nordic Conference on Human-Computer Interaction 2008. pp. 439-442*

*Chapman, Christopher N. and Milham, Russell P. (2006): The Personas' New Clothes: Methodological and Practical Arguments against a Popular Method. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 50 (5) pp. 634-636*

*Chapman, Christopher N, Love, Edwin, Milham, Russell P, ElRif, Paul and Alford, James L. (2008): Quantitative Evaluation of Personas as Information. In Human Factors and Ergonomics Society Annual Meeting Proceedings, 52 (16) pp. 1107-1111*

*Chatman, Seymour (1990): Coming to Terms: The Rhetoric of Narrative in Fiction and Film. Cornell University Press*

*Cobley, Paul (2001): Narrative. Routledge*

*Cooper, Alan (1999): The Inmates Are Running the Asylum: Why High Tech Products Drive Us Crazy and How to Restore the Sanity. Sams*

*Cooper, Alan, Reimann, Robert and Cronin, David (2007): About Face 3: The Essentials of Interaction Design. Wiley*

*Djajadiningrat, J. P., Gaver, William W. and Fres, J. W. (2000): Interaction Relabelling and Extreme Characters: Methods for Exploring Aesthetic Interactions. In: Proceedings of DIS00: Designing Interactive Systems: Processes, Practices, Methods, & Techniques 2000. pp. 66-71*

*Drego, Vidya L. and Dorsey, Moira (2010). The ROI of Personas. Forrester Research http://www.forrester.com/The+ROI+Of+Personas/fulltext/-/E-RES55359*

*Egri, Lajos (1960): The Art Of Dramatic Writing: Its Basis in the Creative Interpretation of*

*Human Motives. Simon and Schuster*

Erickson, Thomas (1995): *Notes on Design Practice: Stories and Prototypes as Catalysts for Communication. In:* Carroll, John M. *(ed.). "Scenario-Based Design: Envisioning Work and Technology in System Development". John Wiley and Sons*

Floyd, Ingbert, Jones, M. Cameron and Twidale, Michael *(2008): Resolving Incommensurable Debates: a Preliminary Identification of Persona Kinds, Attributes, and Characteristics. In* Artifact*, 2 (1) pp. 12-26*

Gazan, Rich *(2005): Imposing structures: Narrative analysis and the design of information systems. In* Library Information Science Research*, 27 (3) pp. 346-362*

Goodwin, Kim *(ed.) (2009): Designing for the Digital Age: How to Create Human-Centered Products and Services. Wiley*

Goodwin, Kim *(2007). Perfecting your personas. Retrieved 15 May 2007 from The Cooper Journal:* http://www.cooper.com/journal/2001/08/perfecting_your_personas.html

Goodwin, Kim *(2008). Getting from Research to Personas: Harnessing the Power of Data. Retrieved 15 May 2008 from The Cooper Journal:* http://www.cooper.com/journal/2002/11/getting_from_research_to_perso.html/

Gothelf, Jeff *(2012). Using Proto-Personas for Executive Alignment. Retrieved 1 May 2012 from Ux Magazine:* http://uxmag.com/articles/using-proto-personas-for-executive-alignment

Grudin, Jonathan and Pruitt, John *(2002): Personas , Participatory Design and Product Development : An Infrastructure for Engagement. In* Design*, 2002*

Hinton, Perry R. *(2000): Stereotypes, Cognition and Culture. Psychology Press*

Jarke, Matthias *(1999): Scenarios for modeling. In* Communications of the ACM*, 42 (1) pp. 47-48*

Kvale, Steinar *(1997): InterViews: An Introduction to Qualitative Research Interviewing. Sage Publications, Inc*

Kyng, Morten *(1992): Scenario? Guilty!. In* ACM SIGCHI Bulletin*, 24 (4) pp. 8-9*

Laurel, Brenda K. *(1991): Computers as Theatre. Reading, MA, Addison-Wesley Publishing*

Long, Frank *(2009): Real or Imaginary: The effectiveness of using personas in product design. In* IES Conference 2009*, (2007) pp. 1-10*

Ma, Jaio and LeRouge, Cindy *(2007): Introducing User Profiles and Personas into Informa-*

*tion Systems Development. In: AMCIS 2007 PROCEEDINGS 2007.*

*Macrae, C. N. and Bodenhausen, G. V. (2001): Social cognition: categorical person perception. In British journal of psychology London England 1953, 92 (0) pp. 239-255*

*Madsen, Sabine and Nielsen, Lene (2009): Using Storytelling to Improve Scenarios. In: Nunes, Miguel Baptista, Isaías, Pedro and Powell, Philip (eds.) IADIS International Conference Information Systems 2009.*

*Mandler, Jean M. (1983): Stories: The Function of Structure. In: Annual Convention of the American Psychological Association 1983.*

*Miaskiewicz, Tomasz and Kozar, Kenneth A. (2011): Personas and user-centered design: How can personas benefit product design processes?. In Design Studies, 32 (5) pp. 417-430*

*Mikkelson, Norunn and Lee, Wai On (2000): Incorporating user archetypes into scenario-based design. In: Proceedings of the Ninth Annual Conference UPA 2000 2000.*

*Mulder, Steve and Yaar, Ziv (2006): The User Is Always Right: A Practical Guide to Creating and Using Personas for the Web. New Riders Press*

*Nardi, Bonnie A. (1996): Studying context: A comparison of activity theory, situated action models, and distributed cognition. In: Nardi, Bonnie A. (ed.). "Context and Consciousness: Activity Theory and Human-Computer Interaction". MIT Press*

*Nielsen, L. (2003): Constructing the user. In: Proceedings of the Tenth International Conference on Human-Computer Interaction 2003. pp. 430-434*

*Nielsen, Lene (2012): Personas - User Focused Design. Springer*

*Nielsen, Lene (2011): Persona: Brugerfokuseret design. Aarhus Universitetsforlag*

*Nielsen, Lene (2004). Engaging Personas and Narrative Scenarios - – a study on how a user-centered approach influenced the perception of the design process in the e-business group at AstraZeneca - PhD thesis. Copenhagen Business School*

*Nielsen, Lene and Madsen, Sabine (2006): Using Storytelling to Reflect on IT Projects. In Journal of Information Technology Theory and Application (JITTA), 7 (4)*

*Norman, Donald A. (2004). Ad–Hoc personas & Empathic Focus. Retrieved 1 January 2004 from JND.org: http://www.jnd.org/dn.mss/personas/empath.html*

*Portigal, Steve (2008): Persona non grata. In Interactions, 15 (1) pp. 72-73*

*Pruitt, John and Adlin, Tamara (2006): The Persona Lifecycle: Keeping People in Mind Throughout Product Design (Interactive Technologies). Morgan Kaufmann*

*Quesenbery, Whitney (2006): Storytelling and Narrative. In: Pruitt, John and Adlin, Tamara (eds.). "The Persona Lifecycle: Keeping People in Mind Throughout Product Design (Interactive Technologies)". Morgan Kaufmannpp. 520-555*

*Ryan, Marie-Laure (2003): Narrative as Virtual Reality: Immersion and Interactivity in Literature and Electronic Media. The Johns Hopkins University Press*

*Rönkkö, Kari, Hellman, Mats, Kilander, Britta and Dittrich, Yvonne (2004): Personas is not applicable: local remedies interpreted in a wider context. In: Clement, Andrew and Besselaar, Peter Van den (eds.) PDC 2004 - Proceedings of the Eighth Conference on Participatory Design July 27-31, 2004, Toronto, Ontario, Canada. pp. 112-120*

*Smith, Murray (1995): Engaging Characters: Fiction, Emotion, and the Cinema. Oxford University Press, USA*

*Sutcliffe, Alistair G. (2003): Scenarios, Models and the Design Process in Software Engineering and Interactive Systems Design. In: Proceedings of the Tenth International Conference on Human-Computer Interaction 2003. pp. 579-583*

*Sønderstrup-Andersen, Eva (2007): Personas. En domæneanalytisk tilgang. In Dansk Biblioteksforskning, 3 (2)*

## About the author

### Lene Nielsen

Lene Nielsen is Denmark's leading specialist in personas. She has a Ph.D.in personas and scenarios and has more than 5 years experience on practical use of personas working with both industry and governmental bodies.

## Your notes and thoughts on chapter 30

Record your notes and thoughts on this chapter. If you want to share these thoughts with others online, go to the bottom of the page at:
http://www.interaction-design.org/encyclopedia/personas.html

**NOTES:**

# CHAPTER 31

# Ethnography

*by Dave Randall and Mark Rouncefield.*

The concern to balance detailed documentation of events with insights into the meaning of those events is the enduring hallmark of ethnography. (Fielding 1994: 154)

[The] immortal ordinary society ... is only discoverable. It is not imaginable. It cannot be imagined but is only actually found out, and just in any actual case. The way it is done is everything it can consist of and imagined descriptions cannot capture this detail. (Garfinkel 1996: 7- 8)

In this chapter we attempt to describe ethnography, its evolution, and how it has been used in human computer interaction (HCI) and computer supported cooperative work (CSCW) research. We begin by discussing ethnography in general and its use in design before going on to focus on one particular variant of ethnography — ethnomethodologically informed, or inspired, ethnography — that has become commonly used as a method (though not as an analytic approach) in

CSCW/HCI research. We conclude by considering some recent developments in ethnographic techniques — especially with regard to 'auto-ethnography' — and a range of problems and complexities in the use of the method in HCI that have arisen in recent years.

## 31.1 WHAT IS ETHNOGRAPHY?

Ethnography is a qualitative orientation to research that emphasises the detailed observation of people in naturally occurring settings. The ethnographic approaches currently used in HCI clearly have their origins in social anthropology. The move towards naturalistic observational methods in anthropology is generally attributed to Malinowski and popularised by other anthropologists such as Boas, and, more controversially perhaps, Margaret Mead (see Freeman 1999; Shankman 2000). These early anthropologists were convinced that only through living with and experiencing 'native' life could a researcher really understand that culture and that way of life, changing the perception of anthropology from being mere 'strange tales of faraway places.' Ethnography also has carved a place within sociology (e.g., the Chicago School. See Hammersley 1990), though it has often been presented as a methodology of last resort — used for obtaining information about deviant groups and cultures — sometimes characterised as 'nuts, sluts and perverts' — that are impossible to investigate in other ways. It has been put to the service of any amount of theoretical work, including feminism, Marxism, actor network theory, activity theory, distributed cognition, symbolic interactionism, grounded theory, and so on ad nauseam. In addition, of course, although ethnography proper is associated with anthropology and sociology, 'fieldwork' can be traced just as easily through cognitive science, Swedish and German work science, and so on.

Figure 31.1: 'Going native', as the ultimate form of ethnography.

*Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.*

Video 31.1: This is the caption text.

*Courtesy of Paul Dourish. Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.View full screen or download (0)*

## 31.2 WHY USE ETHNOGRAPHY?

Perhaps the main virtue of ethnography is its ability to make visible the 'real world' sociality of a setting through detailed descriptions of the 'workaday' activities of social actors within specific contexts. Ethnography seeks to present a portrait of life as seen and understood by those who live and work within the domain concerned, what it terms an 'appreciative stance', through the direct involvement of the researcher in the setting under investigation. It is, as Fielding suggests,

> a stance which emphasized seeing things from the perspective of those studied before stepping back to make a more detached assessment. .... mindful of the Native American adage that one should 'never criticize a man until you have walked a mile in his moccasins.' (Fielding, 1994: 156)

(This, of course, has the added advantage (as the old joke goes) that when you do eventually come to speak your mind, you're a mile away ...and you've got his shoes...)

The intention of ethnography is to see activities as social actions embedded within a socially organised domain and accomplished in and through the day-to-day activities of participants. It is the ability of ethnography to understand a social setting as perceived by its participants (for HCI, the archetypal users) that underpins its appeal. Its chief characteristic, supposedly, is the researcher's (prolonged) immersion in the setting, and the detailed observation of circumstances, practices, conversations, and activities that comprise its 'real world' character. Having said that, and as pointed out by Randall et al. (2007), issues surrounding prolongation, detail, etc. are very much determined in practice by what it is that is being studied.

## 31.3 'DOING' ETHNOGRAPHY — RELYING ON THE 'KINDNESS OF STRANGERS'.

The aim of ethnography is to assemble an account of the way in which people manage and organise their lives, by trying to obtain an 'insider's' view. This necessitates the fieldworker becoming involved in the setting and the activities being studied, counteracting the temptation, when studying others' lives, to simply read things into them — or 'make stuff up'. This is why ethnographic investigation generally insists on approaching the investigation of a setting without any *theoretical* preconceptions as to what will be found, since, much to their frustration, the social world is generally not organised in ways that analysts and researchers want to find it. Moreover, things that are familiar are extremely difficult to see clearly because of their very familiarity.

In terms of the practicalities of ethnographic work, Evans-Pritchard, the famous anthropologist, wrote of how he sought some insight on how to do fieldwork from other noted anthropologists and received advice that amounted to little more than 'don't drink the water and leave the women alone'. While this still probably remains very good advice, it should also be understood that ethnography is neither an esoteric procedure requiring immense amounts of training, nor is it searching for things that are hard to find. Nor, however, is it simply 'hanging around' — or as Button and King (1992) put it, "hanging around is not the point." While much of ethnography does involve 'hanging around', this is not its point but a means of achieving the objective of uncovering the sociality of work. Much of ethnographic practice is simply about presenting oneself as a reasonable, courteous, and unthreatening human being who is interested in what people do and then shutting up, watching, and listening. Ethnography requires simple abilities, including an ability to listen, show an interest in what people do and what they have to say, and tolerate long periods of boredom. Ethnography is an immensely ordinary activity requiring ordinary, mundane skills.

The important thing about the ethnographer is not that he or she brings particularly arcane skills to the collection of data [many of those are the skills of office administration, cataloguing and classifying documents and records], but that they bring the *willingness* to pay attention to people's activities, to attend in detail to how people actually go about their affairs, however ordinary and otherwise unremarkable these affairs might be. (Hughes and Sharrock 2002: 20)

## 31.4 WHAT DOES AN ETHNOGRAPHER DO?

What an ethnographer does is what any other person in the organisation being studied is likely to do — watching, talking, sitting in meetings, learning their way around the organisation. And it is not difficult. The data is not hard to find, the fieldworker does not need to *look* for it, it is right there in front of him or her. And as Sacks (1984) puts it, “there is order at all points.” Consequently, there is no particular need to suffer the fieldwork agonies so well described by Agar in his study of ‘the professional stranger’:

You arrive, tape recorder in hand, with a grin rigidly planted on your face. You probably realise that you have no idea how your grin is being interpreted, so you stop and nervously attempt a relaxed pose. Then you realise you have no idea how that is being interpreted. Soon you work yourself into the paralysis of the psychiatrist in the strip joint - she knows she can’t react, but she knows she can’t not react. It is little wonder that sometimes people hide in a hotel room and read mysteries. (Sacks, 1980: 15)

For most fieldworkers — for us — these agonies, if they occur at all, are rare and short-lived, soon to be replaced by the very different agony of the ‘fieldwork junkie’. Most ethnographers will soon realise that ‘becoming an ethnographer’ has some interesting parallels with Becker’s (1953) analysis of ‘becoming a marijuana user’, such that Beckers’ ideas of ‘learning to recognise’ and ‘learning to appreciate’ resonates with the experience of fieldwork.

In terms of how to behave, while a researcher cannot cope with every personal idiosyncracy, there are some common sense principles of conduct for the ethnographer. These principles primarily involve recognising that for those in the setting, their commitment to what goes on there is their business, their job — and the fieldworker, no matter what his or her personal inclinations are, must respect this. The point of fieldwork is to understand the social organisation of activities within the setting. This requires stringent attentiveness to what persons have to say and do, for the ethnographer, like Blanche DuBois, is generally reliant on the ‘kindness of strangers’. While this does not require an exaggerated show of interest in the (often) boring details of what people do — and most working environments can turn out to be boring places — it does require avoiding prejudgements about what is of interest and what is not.

The ethnographer accesses ‘what is going on’ in a setting through the mundane competences he or she has developed that routinely make it possible to learn about new cultures and forms of social organisation. The apparent ‘strangeness’ or initial unfamiliarity of a field site has an analytic utility in helping the ethnographer reveal and document the methods by which members ‘just do it’ when it comes to everyday, mundane work. The initial strangeness of a setting is consequently regarded as facilitating the necessary distance required to ‘make the ordinary extra-ordinary’ enabling the ethnographer to render the familiar strange yet recognisable.

## 31.5 COLLECTING DATA

In terms of what the fieldworker collects by way of data, experience shows that this is the least of the problems of ethnography, and anyway it will be dictated not by strategic methodological considerations, but by the flow of activity within the social setting. The ‘data’ is often lying around in plain sight, but no one has bothered to collect it up. There is nothing special to look for, nothing to find that is hidden. Hughes and Sharrock suggest that,

another simple truth about ethnography is that, given access, you can very quickly collect far more data than you can ever possibly use: a day’s work can generate several hours of audio or video tape recording. Nor is there really much meaning to the idea that some things are *crucial* data — ethnography is a pretty diffuse exercise with [characteristically] vague objectives, if indeed, they can be called objectives at all: often the aim is just to see and hear as much as you can, and to get as good a record of what you can see and hear as possible. In the ethnographic setting it is *all* data, though there is no sense to having *all* the data. (Hughes and Sharrock, 2002: 20)

The ethnographer's job is to listen to the talk, watch what happens, see what people do, to write it down, tape it, record what documents can be recorded, and so on. The sorts of things that can be collected and recorded include: conversations, descriptions of activities, diagrams of places, job descriptions, memos, notices, graffiti, transcripts of meetings, war stories, and more. It is not that such materials have any intrinsic value; the material is valuable insofar as it can be made relevant or useful for what it can say about the social organisation of activities. Marilyn Strathern (2003) suggests that ethnography is "the deliberate attempt to generate more data than the investigator is aware of at the time of collection" (quoted in Dourish (n.d.: 2), but whilst we share the sense of the mass accumulation of data that often accompanies ethnographic work, there is often little 'deliberate' about the process — it just tends to happen as a consequence of immersion in the setting, everyday curiosity, and the usual researcher anxieties that manifest themselves in 'if in doubt collect stuff'.

## 31.6 ETHNOGRAPHIC ANALYSIS

Almost any fool can collect data — it's not difficult to do. The hard task is to analyse the mass of material and to find out what it all amounts to. This, evidently, very much depends on what you are there to do, who has asked you to do it, and what expectations there might be in relation to output (this is arguably more relevant in interdisciplinary contexts such as CSCW and HCI than it is in more 'purely 'sociological work.) For us, the following precepts have been useful 'aids to a sluggish imagination'.

Precept 1: Assume that the world is socially organised — and show how this orderliness is accomplished in the setting.

Precept 2: See the setting and its activities as socially organised from within — assume that the setting and its activities make sense to the participants and uncover and explicate that understanding.

Precept 3: Understand the setting and its activities in terms that members' understand and use — look at the actual activities as they actually occur during the course of the work.

Precept 4: Examine activities in all their detail.

Precept 5: Treat activities as situated — activities are not isolated events but situated within a context that informs their sense and their character.

Precept 6: Attend to the 'working division of labour' — although individuals perform activities, these are often embedded in interaction and cooperation with others. Understanding how this moment-to-moment coordination achieved is one of the tasks of analysis.

Precept 7: Tasks and activities are sequenced — our activities are, typically, sequenced if only in the highly general way that activities follow one another in some series. Thus, we get up in the morning, brush our teeth, have breakfast, get ready for work, go to work, etc. However, in the case of many activities, this sequencing has strong implications in that the sequencing is integral to the interactional sense of some activity.

Precept 8: Attend to the egological organisation of activities — it is people who do things, not organisations. Actual work is performed by a person who has to determine how his or her activities fit into his or her responsibilities and relevances, and how this will fit with that of others.

Precept 9: Don't draw a distinction between expert knowledge and practical knowledge — avoid the tendency to underrate the skills and competencies involved in even the most routine of tasks, since ‘routineness’ is very often the result of the experienced and practised grasp of complex skills.

Precept 10: Don’t treat settings as equivalent — this is a caution against spurious and unwarranted generalisation.

Making sense of the materials collected is, of course, not a matter of making any sense or, worse, trying to find *the* sense of the materials as if they had only one sense. However, ethnographic research is directed toward some research objective. Its purpose is to develop an analysis and an understanding of a setting that has some relevance. While the fieldworker needs to go into a setting with as few conceptions as to what will be found there, this is a posture designed to further a research aim; in this case understanding particular aspects of everyday, routine work.

## 31.7 THE USES OF ETHNOGRAPHY

Given the very varied research objectives that stimulate research, ethnographic methods are utilised, deployed, and adapted in a variety of ways. These ways often depend on very practical or serendipitous aspects of the research process, such as the complexities of obtaining fieldwork access. This variety of uses does not constitute an obvious research typology, such as those that are frequently produced for participant observation studies, for example, the common distinction between overt and covert observation or Gold’s (1958) typology based on various identified relationships between ‘observation’ and ‘participation’. Instead it suggests an orientation to a range of practical factors, such as available time ‘in the field’, and the availability and suitability of existing data. The different uses of ethnography identified by Hughes et al. 1994 include:

Re-examination of previous studies: Here previous studies are re-examined to inform initial thinking.

‘Quick and dirty ‘ or ‘lightweight’ ethnography: Here brief ethnographic studies are undertaken to provide a general but informed sense of the setting.

Concurrent ethnography: This is the idea of an ongoing ethnography that adapts its focus over time. Here, design is influenced by an on-going ethnographic study taking place at the same time as systems development.

Evaluative ethnography: Here, an ethnographic study is undertaken to verify, validate, or evaluate a set of already formulated design decisions.

These categories should not be read as if they were mutually exclusive ways of using ethnography; some of the uses could be, and were, harnessed together and the differences between them should be seen as differences of emphasis rather than as sharp demarcations. Design is a matter of responding to contingencies of various kinds. Design objectives are various, and this will have a bearing on the role of ethnography. In other words, while not necessarily buying into the picture of the design process as a series of discrete, clearly delineated, and phased steps, it undoubtedly has different objectives at different stages and, accordingly, implications for how design needs to be informed by relevant information about the domain.

## 31.8 ETHNOGRAPHY AND DESIGN: 'IMPLICATIONS FOR DESIGN'

The value of ethnography in design is a matter of controversy (cf. Anderson 1994; Plowman et al. 1995) since there are no panaceas for the problems of design, and arguably could not be. This would entail 'design' having a universal character — which it self-evidently does not — and an entirely predictable problem-solution structure, which it evidently does not, and that is why we distinguish design from IKEA furniture assembly. We can only expect ethnography (or the sociology that may be associated with it) to have a modest utility to design, and the role of ethnography as we practise it is primarily as an 'informational input' into design, and, as such, only one source of information. The input can be of *critical* value insofar as it can advise the designer of actual practices of work and may clarify the role that actual practices play in the management of work; matters that may not normally be captured by other methods. In as much as a position on the role of ethnography in CSCW design has emerged, it can be expressed in its ability to make visible the everyday nature of work. As Suchman writes,

ethnographies provide both general frameworks and specific analyses of relations among work, technology and organisation. Workplace ethnographies have identified new orientations for design: for example, the creation and use of shared artifacts and the structuring of communicative practices. (Suchman 1995: 61)

This is, in fact, a 'sociologically partisan' conception of ethnography, but it does have the advantage of focusing upon the specific and detailed organisation of activities and, thereby, upon the very activities which designers are concerned to understand, analyse, and reconstruct. It is the ability of ethnography to describe a social setting as it is perceived by those involved in the setting, (the archetypal 'users'), that underpins its appeal to designers. In particular, it offers the opportunity to reveal needs or practices of users which they may not themselves attend to because they take them so much for granted that they do not think about them. In other words, we are dealing with 'needs' which they cannot articulate because of the bureaucratic or power relationships within which they are placed or because they are simply too busy. As part of the initial process of requirements capture, ethnography is valuable in identifying the exceptions, contradictions, and contingencies of work activities which are real conditions of the work's conduct, but which will not (usually) figure in official or formal representations of that work.

The assumption is that it is for *designers* to draw design conclusions from the results of ethnography. The kinds of changes to design that will result from this approach are intended to have an *incremental* rather than a *comprehensively transformative* effect. There is no intrinsic design significance to the results of an ethnographic study, for such significance must be relative to the nature of the design exercise itself, to the purposes, conceptions, methods, and plans of those making the design. Ethnography should be done independently of design preconceptions, distancing itself from the preoccupations, enthusiasms, and orientations of the designer, and refraining from looking at the setting and its affairs 'through designer's eyes'. While there may be a tension between the designer's and the fieldworker's roles, this is a positive feature, something that is hardly likely to be destructive of good design, through highlighting the difference between good *abstract* design solutions, good *practical* design, and, ultimately the social and political effects of design solutions. (Dourish 1996). In this way, to paraphrase the sociologist Max Weber, we may think of ethnography as being 'design relevant' but not 'design laden'.

What seems to be a largely commonplace observation like this has proven controversial. In particular, the relationship between ethnography and design was subjected to a forensic lens by Dourish in his well-known paper, "Implications for Design" and has been robustly criticized by Crabtree et al. (2009) ('Ethnography Considered Harmful'). It is worth examining this argument. For Dourish, the relationship between ethnography and design has been under-examined. There are two consequences of this. Firstly, it has led to some naïve renderings of design implications towards the end of otherwise competent ethnographies; secondly, and this is a slightly different argument, it has led to the naïve acceptance of what we will call a 'service' relationship which ignores the potential that ethnography has for a more critical — perhaps overtly political — role. We share Dourish's view of the naïve service relationship and regard a preference for critique and political intervention benignly as well (Howard Becker's famous 1967 paper "Whose side are we on?" describes this as well as anything we have read). However, Crabtree et al.'s sometimes misunderstood position does not run counter to this. We have been at pains to emphasise ethnomethodology's rejection of analytic 'privilege'— that it cannot claim to provide accounts that are 'superior' in virtue of the professional status of practitioners. We can only claim that we do solid, detailed, empirical work that others may not be minded to do for a variety of reasons. The issue in respect of critique is, for us, whether there is any reason to believe that a professional social scientist offers better critique than anyone else. We do not think so, for to adjudicate such matters would require us to adjudicate in the first place what the grounds for critique might be, and it is precisely the case that those who disagree with us may well disagree as to what those grounds should be. Crabtree et al.'s argument is broadly predicated on the view that such accounts are not especially useful (to design), although they do say, 'we do not dispute the need for critical reflection in design or any other technical practice as that notion is ordinarily understood.' (Crabtree et al (2009): 884) They mean, however, that people other than social scientists are perfectly capable of taking a critical view. Such an argument is and always has been deeply unpopular with professional practitioners of the social sciences. To put it another way, the proper relationship for Crabtree et al. is a relationship between data and design, and good data is obtained by ethnomethodologists. For Dourish, the issue is less about data than it is about the way in which data *is cast* so as to serve distinctive and critical purposes. Our view, for what it is worth, is that no strong relationship between ethnography of whatever kind and design has ever been established in the workplace or elsewhere for the simple reason that this relationship is always and everywhere contingent. Other renditions of this relationship can be found in papers such as Button and Dourish (1996); Button and Dourish (1998).

Having said all this, there must be some purpose to ethnographic enquiry in HCI and CSCW for, if not, why do it? Ethnography originally became popular in HCI and CSCW in the 1980s and 1990s because of its claim to provide a method more attuned to the socially organised character of workplace settings. This 'turn to the social' in design and the interest in ethnography arose out of dissatisfaction with existing methods of informing design as offering overly abstract and simplistic analyses of social life. Ethnography with its emphasis on the in situ observation of interactions within their natural settings seemed eminently suited to bringing a social perspective to bear on system design. The 'turn to the social' recognised a new kind of end-user, a 'realtime, real-world' human being, and consequently designers turned to the social sciences to provide them with some insights, some sensitivities, to inform design. The advantage of using ethnographic methods in CSCW for studying work lies in the way it documents the real-world character and context of work and the opportunity it provides to ensure system design resonates with the circumstances of its use. In attempting to document, describe, and account for activities, ethnography seeks to provide an answer to what might be regarded as the essential CSCW and design question (Shapiro 1994), 'what to automate and what to leave to human skill and experience'.

Even though newer approaches, such as 'cultural probes' have since then come along, ethnography has remained surprisingly popular, long past the initial enthusiasm that often accompanies any new approach, to the extent that some form of ethnography or ethnographic study sometimes seems a necessary first step in any HCI/CSCW investigation. If design, as a 'satisficing activity', is more of an art than a science, dealing with messy indeterminate situations and 'wicked problems' (Rittel and Webber 1973), then before designers can solve a design problem, they need to understand some basics, such as what they are designing, what it should do, and who should use it and in what circumstances. It was argued that ethnography was the method attuned to gathering exactly this kind of relevant data. That is, there are certain kinds of things that ethnography might normally be said to provide:

Additional domain knowledge.

An overall view of complex settings which would otherwise be difficult to obtain.

Perspectives from, and practices of, a variety of stakeholders.

Some assessment of the scope and limitations of systems and products that might be envisaged.

A balanced view of the relationship between standardised processes, human skills, and how to deal with contingencies.

A fuller view of the real-world nature of the problems that need to be solved.

Detailed knowledge of the routine ways in which technologies actually get used, and what for.

A critique of 'snake oil' salesmen — i.e., those who offer simplistic technical or organisational solutions.

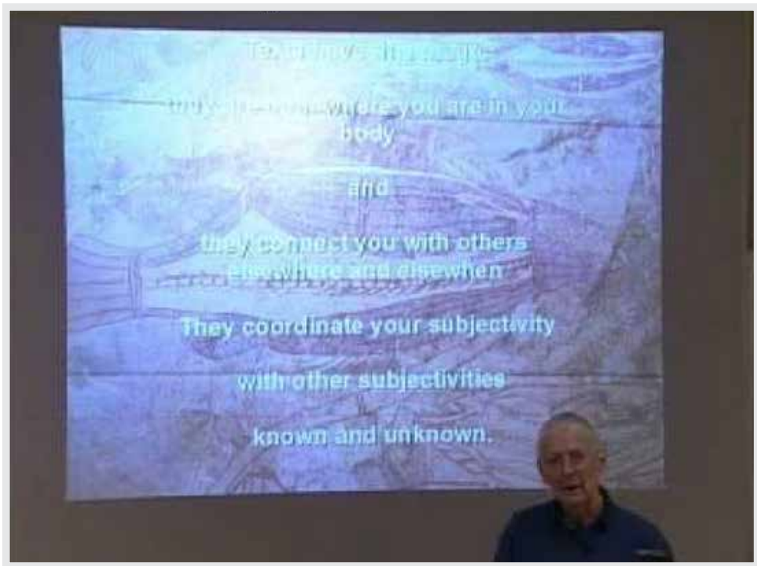
However, the relation between ethnography and any design ambitions has always been somewhat problematic (see for example Plowman et al. 1995). As with all radical changes in perspective, initial enthusiasm has been followed by rather more critical reflection. While ethnography may have been effective in providing a critique of systems design, it has been less adept at producing design solutions and translating ethnographic insight into good design practice. Some would argue that simply documenting and describing the grossly observable features of a setting — termed 'scenic ethnography' (Button 2000) — does not do much neither to inform us about the processual and interactive features of a setting, nor to provide design recommendations (see also Crabtree et al. ' 2009). The more cynical, amongst us (and we should probably include ourselves in that number) would suggest that simply going out and doing some observations is no panacea for the problems of design — as we have said already, but it's worth repeating (on the principle that if something is worth saying it's probably worth saying twice), there really is no silver bullet. Those researchers who have carried out ethnographic studies have long been aware of its limitations when it came to translating ethnographic findings into design recommendations or requirements and have responded to this challenge in various ways. Some have provided a series of tenets to guide the ethnographer (Sommerville et al. 1992) to look, for example, for those aspects of a setting's organization that need to be retained in any work redesign. Others (Hughes et al. 1995;Hughes et al. 1997) have talked about the value of ethnography in providing ‘sensitivities’ to designers — and particularly in providing some clues as to what designers should not do. Others again have seen the problem in terms of the way in which ethnographic findings have been reported or represented and have produced various approaches — such as ‘the Designers’ Notepad’ (Hughes et al. 1995) or the use of ‘patterns’ (see Martin et al. 2002) — that aim to make the long discursive texts typically produced by ethnographers rather more ‘designer friendly’.

Any number of idiocies (and it’s definitely more than one) have emanated from commentators wishing to discuss the relationship between ethnographic data and the design process. Not least, one might imagine that some examination of what kinds of design, done by what kinds of designer, in what kinds of organizational (or other) context might be conducted before we make crass judgments about this relationship. After all, at the outset, the problem space that ethnographies were intended to address was quite narrowly defined — studies of work and organization designed to aid the design of collaborative computer systems. That is no longer true. Even a moment’s thought tells us that it is absurd to hold ethnography to account for design decisions if the design space is now so vast. It is hard to think of any human (or other) context that cannot be designed for. After all, we were involved in evaluative work where we discovered that one of the main uses for a camera technology (intended to be a memory aid) was held by users to be an opportunity to see what life looked like for cats, dogs, and children (Harper et al. 2007). Such decisions are contingent, and may well be out of the hands of both ethnographers and people who might normally be thought of as the designers. Even beginning to get to grips with this issue requires us to confront some intractable problems, and so there are no general solutions to the problem of relating ethnographic enquiry to design — there are only specific problems. It will depend on the many and varied possible uses to which ethnography can be put, the kinds of design team in which the data are to be examined and used, the scale of the project in question, the relationship of ethnography to other methodologies which might be in use, and so on.

In our view, the professional demands of ethnography are exaggerated, and we are equally negative about the way in which ethnography is viewed, more or less unproblematically, as an alternative (sometimes the only) method. Rather, it is a tool in the toolbox — not only for designers but for anyone who wants to know what needs to be changed and how to go about changing it. Ethnography is always about asking questions such as, 'What kind of problem have we got?' What does the problem look like? How does it manifest itself?' before beginning to provide design solutions to the problems we identify. Similarly, at the same time, an interdisciplinary sensitivity requires us to take design seriously, understanding how designers go about solving their problems, identifying candidate solutions, and applying their technical knowledge to them. What any ethnographer (ethnomethodological or otherwise) seeks to do is establish what questions seem relevant and what might be the best ways of getting robust and reliable answers to those questions.

Video 31.2: Have another look at Dourish - the implications of anthropological work for design.

*Courtesy of Paul Dourish. Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms. View full screen or download*




Video 31.3: Dorothy Smith on Institutional Ethnography.

*Courtesy of Dorothy Smith. Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms. View full screen or download (0)*

## 31.9 ETHNOMETHODOLOGICALLY INFORMED ETHNOGRAPHY: 'WELCOME TO THE DARK SIDE'.

I want to encourage the sense that interesting aspects of the world, that are as yet unknown, are accessible to observation. (Sacks 1992: 420)

... [The] immortal ordinary society ... is only discoverable. It is not imaginable. It cannot be imagined but is only actually found out, and just in any actual case. The way it is done is everything it can consist of and imagined descriptions cannot capture this detail. (Garfinkel 1996: 7- 8)

Video 31.4: Ethnomethodology by Wes Sharrock.



Ethnography is not in any sense a unitary method, but is a gloss on various and different analytic frameworks — thus there are, as we have intimated, Marxist, feminist, and postmodern ethnographies. Here, however, we provide some detail of one type of ethnography — ethnomethodologically informed ethnography — and how it can be deployed to discover some of the features of 'immortal ordinary society' or everyday work and life, and then we consider some aspects of the analytic purchase this approach brings to the understanding of 'real time real world work'. This emphasis on 'real world, real time work' stands in rather stark contrast to many sociological accounts of social life in general and (perhaps) the everyday world of work in particular. Conventional sociological accounts portray a world in which not only does "homo sociologicus' neither laugh nor cry' (Williamson 1989), but does not seem to do much that looks like work either. This appears to be a world in which the practical accomplishment of work — the skills, and competencies that workers routinely and visibly bring to their jobs — is largely absent. Consequently, although there are many sociological studies of 'work', they often seem to have very little to say about the actual work which goes on within the setting under study — about what makes *this* work 'bank work' or 'insurance work'. In the process, both the worker and the fashion in which work is accomplished effectively disappear into theoretical abstraction. The desire to be attentive to the work is, therefore, one of the motivations for the use of ethnomethodologically informed ethnography. In contrast to a common sociological attitude which views specific social settings as sites of generic, abstract social processes, the ethnomethodologically informed ethnographic approach is particularly focused upon the distinctiveness and the specificity of the settings under study.

There have been a number of attempts to document the characteristics of 'ethnomethodological ethnography' (Dingwall 1981) or 'ethnomethodologically inspired ethnography' (Silverman 1985). Dingwall, for example, outlines the following characteristics: accomplishing social order; specifying actors' models; suspending a moral stance; creating 'anthropological strangeness'; and depicting stocks of knowledge.

Emerson and Pollner, however, argue that:

the overlap of genealogies, concerns, and prefixes might lead one to expect a cordial relationship between ethnomethodology and ethnography... both perspectives are informed by the interpretive tradition, concerned with the lifeworld, respect the point of view of the social actor (hence 'ethno-'), and typically eschew quantitative and theoretical approaches ... despite the similarities, however, the relation has not been congenial...(Emerson and Pollner, 2001: 118)

They go on to argue how over the years the boundaries between ethnography and ethnomethodology have become blurred, and that recent attempts to integrate ethnomethodology and ethnography (Silverman 1993; Gubrium and Holstein 1997) suggest that once pronounced differences may be dissolving into an integrated methodological sensibility. They also say, however, that ethnomethodology challenges key aspects of ethnographic theory and practice, and that it faults ethnography

for being both too involved in and too removed from the social worlds it studies, and for ignoring the problematics of its own efforts to represent such worlds" and that "self-deconstructing aspects of EM provide good reasons for EG not to embrace EM initiatives too enthusiastically .. EM insights can be used selectively to heighten sensitivity to fundamental methodological issues and to augment appreciation of the practices of both subjects of ethnography and ethnographers themselves. (Emerson and Pollner, 2001: 118)

This section is primarily concerned with documenting the 'analytic purchase' of ethnomethodologically informed ethnography and, in consequence, its utility for describing and understanding everyday organisational activity. While an ethnographic stance arguably entails some minimum orientation of viewing the social world from the standpoint of its participants, one approach to this is the ethnomethodological one, in which members' methods for accomplishing situations in and through the use of local rationalities become the topic of enquiry. For ethnomethodologically informed ethnographic enquiry, members and their subjective orientations and experiences are central. Observation focuses on the places and circumstances where meanings and courses of action are constructed, maintained, used, and negotiated.

Their rational features consist of what members do with, what they 'make of' the accounts in the socially organized actual occasions of their use.." (Garfinkel 1967: 2-3 )

In ethnomethodologically informed ethnographic research on work, the understanding of any work setting is derived from the study of that setting itself, rather than from any highly structured model or theory of work organisation or work processes; that is, it ties itself closely to the observed data, it is 'data-driven'. A central precept of ethnomethodological ethnography is to aim to find the orderliness of ordinary activities, an orderliness accomplished by social actors, unreflectively taken-for-granted by them and constructed with their common-sense knowledge of social order.

The purpose of ethnography is then to display the 'real world' social organisation of activities. Ethnographic studies focus on 'real world, real time' activity, following courses of action as they happen. This requires showing not just *that* some setting is socially organised, but to show in detail just *how* it is organised. The relevance of an ethnomethodologically informed perspective lies in the fact that this respecification of sociology draws attention to the way in which orderliness can be viewed, *inter alia*, as a feature of the sense making procedures participants use in the course of their work. In documenting how work is socially organised, research reveals facets of mundane organisation, how, for example, individuals are enabled to work because of their awareness of what constitutes their task and its linkages with other tasks — the 'egological' division of labour.

In acknowledging the 'situated' character of work, ethnography displays how even in the most apparently routine activities workers need to use their judgment and discretion in response to the various contingencies that arise. Furthermore, 'real world, real time' activity is not necessarily confined to the specific, immediate, locally bounded situation. The sense of what a person is doing here and now is dependent on how that activity is situated within a whole set of understandings about organisational processes, institutionalised patterns, and so on. The organisational context, then, is relevant to the work-in-hand, and ethnography's concern with the organisational context of work is a concern for how aspects of the organisation are relevant to and reflected in on-going everyday, routine work. The organisation is relevant to and reflected in the local work situation as a practical consideration. In consequence, the accomplishment of work tasks involves a range of tacit skills and local knowledge that may be rendered invisible by formal models of processes or procedures, often going unrecognised by the workers themselves; skills which may become visible only when routines or organisations break down and fail to deliver.

## 31.10 REAL WORLD, REAL TIME ACTION - BEING 'LEAD BY THE PHENOMENA'

When used from an ethnomethodological stance, ethnographic work involves a renewed and unprejudiced look at the phenomena that have frequently become obscured beneath layers of theoretical abstraction and speculation. It sets out a policy whereby

No inquiries can be excluded no matter where or when they occur, no matter how vast or how trivial their scope, organization, cost, duration, consequences ... (Garfinkel 1967: 32)

The aim is to observe and describe the phenomena of 'everyday life' independently of the preconceptions of conventional sociological theories and methods. In this approach, observations are 'led by the phenomena', rather than by the concerns and requirements of a particular sociological theory. This means, in effect, that one takes an 'unmotivated' approach to the activities, looking just to see what people are doing, rather than seeking to identify things which are sociologically interesting . Ethnography in general recognises a great temptation when studying other people's lives to read things into them, but ethnomethodologically- informed ethnography in particular is predicated on the view that the social world is not always organised in ways that analysts and researchers want to find it, and hence resists imposing a prior analytic framework on the phenomenon.

This involves dispensing with conventional sociological preconceptions that there are numerous things people are doing which are trivial and not worth observing. These things are trivial in a *sociological* sense, i.e., do not matter with respect to the kinds of things sociologists think are important about a given activity. Ethnography does not seek to explain the orderliness of work activities as the result of factors external to that setting, such as 'power', but treats activities as necessary activities-in-a-social-setting proposing that members display an everyday attentiveness to the socially situated character of their own and each other's actions. The mere fact that people are doing it justifies the attention being given to it by an ethnomethodologically informed ethnographer. In this way the 'false starts', 'glitches', 'diversions', 'distractions', 'interruptions', 'digressions', which are aspects of all activities and notable features of the phenomena, not, so to speak, 'noise' to be eliminated from the data in order to reveal 'essential' or 'sociologically relevant' aspects of the data.

VIDEO 31.5: Ethnomethodology: Mike Lynch on Ethnomethodological studies of work in the sciences.

*Courtesy of Mike Lynch. Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms. View full screen or download (0)*

In ethnomethodologically informed ethnography, the phenomena are to be studied in their character as 'phenomena of everyday life', as 'everyday' occurrences for those who are involved in the activities in question, and the investigator is, therefore, seeking to ascertain what the phenomena *mean* for them. Ethnography assumes that the setting and its associated activities make sense to the participants, and the interest is in descriptions of activities as understood by parties to the setting as opposed to analysts' descriptions. It is not for the investigator to decide what things are, what matters, what is important, or trivial, but to ascertain how things are judged in that way by those who are doing them and to examine the familiarity with and understanding of these matters possessed by those who must live with them. In studies of the kind that ethnomethodologically informed ethnographers make, the concern is with the depiction of ‘the working sensibility’ of those under study. The interest is remote from the kinds of general reflections that someone in an occupation can produce, and much more engaged with their consciousness and attention when they are ‘at work’: what kinds of things do they take for granted or presuppose in going about their work; what kinds of things do they routinely notice; what kinds of things are they ‘on the lookout for’; how do they ‘tune themselves in’ to the state of being ‘at work’; what are the constituents of their ‘serious frame of mind’; how do they react to the things that occur within their sphere of attention; what objectives are they seeking to attain in their reactions to whatever occurs; and by what means — through what operations — will they seek to accomplish those objectives in adaptation to these unfolding circumstances. Thus, attention is focused — in a way that is otherwise almost unprecedented in sociological studies — upon the study of *doing the work*. The emphasis is on ‘work in the raw’, work as it is done, and in the ways in which it is done in *actual practice*, as opposed to work in *idealised* form.

## 31.11 OBSERVING FEATURES OF SOCIAL ORGANISATION - ‘PRACTICAL ACTION’

People who are constantly asking ‘why’ are like tourists who stand in front of a building, reading Baedeker, & through reading about the history of the building’s construction etc. etc. are prevented from seeing it. (Wittgenstein 1984: 40)

The features of everyday social organisation that ethnomethodologically informed ethnography brings to the study of work, technology, and organisations would ‘typically’ include some notion of the visibility of social organisation; an explication of the world known in common and the intractable practicality of action. This approach involves attending to the work and the accountable character of work, attempting to take work seriously — that is, as *work* and not as the manifestation of some grander speculative theory.

The ethnomethodologically informed orientation to ethnography begins from the point of view of the social actor acting within a socially organised environment. The presumption of a 'world known in common' is an assumption about the mutual orientation of members of society in the mundane construction of daily life and is treated as a condition of ordinary concerted action. The relevance of this to ethnography is that the multifarious ways in which the world is assumed to be 'known in common' are apt to be taken-for-granted, to be treated as things that are of such patent obviousness and familiarity that they need not be paid direct and explicit attention. But the investigator is not merely seeking to capture the standpoint and experience of the participant in the setting in respect of the things which that participant might note, explicitly comment upon or pay significant attention to; he or she is also looking to identify those things which the participant is not explicitly attending to, but is nevertheless depending upon. These are the features of the organisation of conduct within the setting that are 'seen-but-unnoticed', but which have *presupposed, taken for granted* status.

Ethnomethodologically informed ethnography is the study of people who are engaged in *practical* action. It is assumed that this is the orientation that pervades the world of everyday life. In everyday life, people give priority to getting things done, and their action is, therefore, organised with respect to the necessities of practicality, and they are engaged in doing whatever it takes to get the things done. For that reason, the purpose of observation is to identify the specific activities in which participants engage to deliver some specific end, and the character of those activities is dictated by the 'specificity of the circumstances'. The essence of practical action is the need to do whatever is to be done under *just these* circumstances, and therefore it involves the adaptation of the course of action to the exigencies of its circumstances. Hence, the concern of the ethnographer lies in the *interplay* of standardisation and specificity. The focus is on the way in which those involved in social settings seek to achieve standardisation of ways of acting, (so as to engender articulated and structured procedures for carrying out relevant types of social action), but it must, at the same time, enforce and implement these in contingent, unforeseen circumstances that may be more or less tractable to compliance with those very standardisations. This accounts for the concern with organisational plans and procedures, and with the way in which the 'idealisations' of courses of action and their circumstances must be articulated with 'actualities'. And it engenders the desire to gain (fieldwork) access to the ways in which work is done *in practice*, and motivates the noticing of the ways in which people achieve (or fail to achieve) conduct in accord with the standardisations that they seek to implement. This gives a reason for putting the exigency and variability of practice into a prominent position in fieldwork studies, one which would be lacking from many sociological approaches because those contingencies and variabilities would not, for that approach, be considered sociologically significant.

## 31.12 ATTENDING TO THE LIVED DETAIL OF EVERYDAY WORK

It is every field researcher's experience that their sense of the definite character of the organization of the 'field', and their sense of the activities they witness within it, *develop together over the course of their involvement* in it. Starting out with only vague notions of how 'such places' conduct themselves, and in the sure knowledge that there are many things going on before them now which they cannot adequately comprehend. They develop, over the time of their inquiries, a considerably fuller sense of what the ways of the setting are and of the character of the occasions that they witness, the two going of course, hand-in-hand. (Sharrock and Anderson 1991: 165)

In advancing ethnomethodologically informed ethnography and contrasting it with other and different approaches in sociology, our emphasis is on 'relevance', on why this approach is particularly relevant to informing ethnographic studies of work, technology, and organisations (and pretty much everything else too). Thus, it is a simple fact that many sociological approaches would not be motivated to do ethnographic studies at all, and that others who were motivated to do so would not — for their own good reasons — consider the practicalities of activities worth noticing. Another point of differentiation is that many sociological approaches are inclined to shift attention away from the activities that are the very business of the setting under investigation. As was suggested earlier, the case of studies of work is a leading example, for though there are many sociological studies 'of work', they have very little to say about the work which goes on within the setting under study.

It is a commonplace sociological attitude to view specific social settings as sites of generic, abstract 'social processes' — for example 'social control' or 'domination' or 'surveillance'. Sociology's purpose in surveying actual social settings is consequently to *minimise* the differences between them, to abstract from the data ways that exhibit the commonality of such processes, to make the case that these are generic. The ethnomethodologically informed ethnographic approach, in contrast, is particularly focused upon the distinctiveness and the specificity, of the setting. Though there may be abstract, general similarities between one setting and another, it is unavoidable that one in the organisation of *practical* conduct must come to terms with the particularities of the setting if the day-to-day affairs of the setting are to be carried out. In terms of many sociological strategies for generalisation, the fact that people are engaged in a particular kind of work is only an *analytically incidental* feature of what they are doing. It is only a concrete instantiation of abstract, generic, and formal processes, which means that there is little investigative motivation to attend to the practicalities of work activities and to the nature of those activities as realisations of the kind of work that they are. In contrast, the ethnomethodologically informed approach has every reason to attend to the distinct character of the work in the setting; e.g., to give priority to the fact that these persons are 'authorising a bank loan', or 'completing a standing order'. Ethnomethodologically informed ethnography directs its attentions to the activities which specifically and distinctively comprise those particular types of activity, and, thus, tries to give detailed characterisations of, and to seek to understand the particular circumstantial conditions for, carrying out those activities in actual cases. The relevance of this to our understanding of work, technology, and organisations has been, then, in the engendering of studies directed toward understanding how the work gets done, and thus to describing the detail and intricacies of working practices for their own sake.

## 31.13 EVERYDAY WORK AS ACCOUNTABLE AND COOPERATIVE ACTIVITY

Ethnography is interested to understand how people make sense of mundane activities and how they make those activities 'accountable' to others. For ethnomethodologists, how people go about making sense of the social world represent mechanisms through which social structure is created, ordered, and sustained. As the social order is continually constructed and reconstructed, members, as 'practical sociologists' are involved in a constant, if taken for granted, process of analysis, so that they are able to act successfully in relation to others for everyday practical purposes. Members must be able to make the social organisation of their mundane activities visible, 'accountable', 'observable-reportable' to each other. The methods that members use to make sense of what is going on are publicly available resources for the observer. Consequently, ethnography is particularly attuned to revealing cooperative aspects of working life — how people reconfigure their arrangements in the face of contingencies and circumstances as they arise.

Social activities are 'concerted' activities involving different people — often very many people — fitting their activities together in quite complex patterns *from within* the activity itself. The expression 'accountable character of activities' refers to this process of concerting work to the way in which people engaged in an activity have to organise their own actions in order that other participants can see what they are doing, and can adapt to it. Participants in social actions have, therefore, to 'make visible' the identity of their actions, to enable other people to identify those actions, and to identify also their purposes and intentions in a way that they can response appropriately to them. This enables them to align their own actions, reciprocally, in the activity that they are jointly, collectively, accomplishing. So, for example, a bank robbery is a collective endeavour in which both the robber and the cashier have to recognise, and make recognisable, their respective roles. The notion of the 'accountable character of activities' emphasises the degree to which activities are organised so as to be identified, recognised, and understood as the activities that they are. The important point is that other people can see what is being done and, thus, consider how they can respond appropriately to align their own actions in the unfolding drama.

That social activities are concerted is a commonplace, if not the *raison d'etre*, of sociology. However, the concern to understand *just how* such concerting takes place (as opposed to why such concerting takes place), *how* people manage to make their activities fit together whilst doing those same activities, appears in the province of ethnomethodology. Its concern with the question of how concerted actions are concerted, and the associated emphasis upon the 'accountable character' of work, has combined to give studies a focus upon the ways in which the pattern of complex activities are 'made visible' to those carrying out those activities. And, simultaneously, they focus upon the ways in which people placed within some complex of action can figure out what is happening around them and how they can fit their own activities into that complex; both when the pattern of activity is a localised one, within their visual field, and the participants can directly monitor those activities which are relevant to their decision as to what to do next, and when they are engaged in patterns of distributed activities.

## 31.14 COOPERATIVE WORKING

The emphasis upon the 'accountable character of activities' explains another relevant aspect of this approach to fieldwork, the focus on cooperative work — the concerting and the articulation of activities (work) from within the activities (work). This involves an interest in how work is accomplished under distributed conditions and of the role of 'awareness', which refers to the ways in which workers can attune themselves to the state of the work process, and integrate their own activities — immediately or remotely — with those of other participants in the work process. This explication of sense making machinery has often invoked work activity as a manifest 'working division of labour' (Anderson et al 1989). Ethnography seeks to understand the organisation of work, its flow, and the division of labour from the point of view of those involved in the work. Because work settings are organised around, through, and within a division of labour, work activities are necessarily seen as interdependent. Understanding how members coordinate their work in real time, moment-by-moment and how they orient to the 'working division of labour' to make sense of what they are doing (Anderson et al 1989) is a feature of ethnographic explication. Ethnography approaches the flow of work (rather than the disembodied idealisations of 'workflow') as an accomplishment, a collective achievement. Consequently, it requires examining the *actual* flow of work, not some idealised version of it. Individuals perform *their* tasks within the context of others similarly doing *their* tasks, within sequences of activities, but the actual work requires individuals to determine and dispay how their work fits into their responsibilities, their relevances, and how this will fit with that of others. Anderson et al (1989) call this an ‘egological’ viewpoint; a view of the world of work and its organisation from the perspective of individuals cooperating and coordinating their activities with others. While individual workers have individual tasks to perform, they are also, and necessarily, individuals-as-part-of-a-collectivity, and much of their work consists in the ability to organise the distribution of individual tasks into an ongoing assemblage of activities *within* a ‘working division of labour’. Individuals, that is, orient to their work according to ‘egological’ principles and their own ‘horizons of relevance’ but have to be attentive to the work of others in order to organise the flow of work in a coherent way. This focus has arguably provided an important analytic tool for the examination of work as lived experience, providing important clues as to both how work was accomplished and, perhaps, why work was done the way it was.

## 31.15 ETHNOGRAPHY AND COMPUTER SUPPORTED COOPERATIVE WORK (CSCW)

Within sociology, ethnography has been deployed to study an array of topics. In CSCW it has primarily focused upon the study of work and settings for which new technology is being designed with the intention of informing that design (Hughes, Randall, and Shapiro 1992; Heath and Luff 1992; Suchman 1983). Ethnography, and especially ethnomethodologically informed ethnography, has acquired some prominence (not to say notoriety) in recent years within the study of CSCW. Ethnography has gained some distinction as a fieldwork method that could contribute both to a general understanding of systems in use in a variety of contexts and to the design of distributed and shared systems (Hughes and King 1992). Efforts to incorporate ethnography into the system design process have had much to do with the (unfortunately belated) realisation, mainly among system designers, that the success of design has much to do, though in complex ways, with the social context of system use. A number of well publicised ‘disasters’ (The London Ambulance System, the Taurus System for the Stock Exchange, for example) suggested that traditional methods of requirements elicitation were inadequate, or in need of supplementation, by methods better designed to bring out the socially organised character of work settings.

This 'turn to the social' in design, the interest in the role of social science theories and approaches in informing design, arose out of dissatisfaction with existing methods of informing design as offering overly abstract and simplistic analyses of the nature of social life. If design, as a 'satisficing activity' is more of an art than a science, dealing with messy indeterminate situations and 'wicked problems'; then before designers can solve a design problem, they need to understand some basics — such as what they are designing, what it should do, and who should use it and in what circumstances. It was argued that methods needed to be more attuned to gathering relevant data in 'real world' environments; that is, settings in which systems were likely to be used rather than in laboratories or other artificial and remote environments. The 'turn to the social' recognised a new kind of end-user, a 'real time, real world' human being, and consequently designers turned to the social sciences to provide them with some insights, some sensitivities, to inform design. Ethnography with its emphasis on the *in situ* observation of interactions within their natural settings seemed eminently suited to bringing a social perspective to bear on system design.

With its emphasis on the 'real world' character of work settings, ethnography is often contrasted with what are commonly regarded as unrealistic and unsatisfactory notions about both systems and the users of systems that tend to be proffered by more traditional methods. Traditional methods of system design perhaps owe far too much to the needs of engineering, and, as a consequence, important aspects of the 'real world' of work are obscured, misrepresented or ignored. It is in this respect that 'analytic approaches', Task Analysis, or Office Automation for example, are found wanting (Shapiro 1993; Suchman 1983) representing an intrusion of the 'engineering mentality' into areas where it is inappropriate. The analytic deconstruction of work activities into ever more finely grained components removes the essential 'real world' features which make them practices within a socially organised setting. This complaint attacks the individualistic slant of the cognitivism which underlies ‘analytic approaches’ by acknowledging the implications of the observation that, as already suggested, work is, typically, collaborative. Though performed by individuals, the various activities that constitute work are performed within an organised environment composed of other individuals, and it is this that gives shape to the activities as ‘real world’ activities. Thus, the focus of ethnography is on the social practices that enable the very processes that ‘analytic methods’ identify, but at the same time decontextualise. It is through the social practices that ethnography seeks to identify and describe that work processes are established and are, accordingly, rooted in socially achieved sets of arrangements.

Such an approach also meshes with the growing use of information technologies within working life. As computers increasingly, and seemingly inexorably, are adopted and diffused into the world of work and organisation, there is a growing awareness that the ubiquitous nature of networked and distributed computing poses new problems for design, requiring the development and deployment of methods that analyse the collaborative and social character of work. Systems are used within populated environments that are, whatever ‘technological’ characteristics they may have, ‘social’ in character and thus the intent of CSCW to design distributed and shared systems means that this social dimension has to be taken into account. Requirements elicitation has to be informed by an analysis of the ‘real world’ circumstances of work and its organisation (Goguen 1993). The virtue of ethnographic approaches comes from the ‘grounded’ recognition that computers are enmeshed into a system of working as instruments and incorporated in highly particular ways — used, misused, modified, circumvented, rejected — into the flow of work. One of the virtues of ethnography lies in revealing these myriad usages in the context of ‘real world’ work settings; furthermore being

more capable than most methods of highlighting those ‘human factors’ which most closely pertain to system usage, factors which are not always just about good interface design but include training, ease of use in work, contexts full of contingencies which are not the remit of system design...even though design may be concerned with developing a completely new system, understanding the context, the people, the skills they possess are all important matters for designers to reflect upon... (Button and King 1992)

The advantages of using ethnographic methods in CSCW for studying work lie in the 'sensitising' it promotes to the real world character and context of work, i.e., in the opportunity it provides to ensure system design resonates with the circumstances of its use. In attempting not only to document or describe activities but also in accounting for them, ethnography seeks to answer what might be regarded as an essential CSCW question as to what to automate and what to leave to human skill and experience. Ethnographic methods thereby assist in the delineation of work design 'problems' as a consequence of greater knowledge of the social organisation of work — the recognition that 'problems' need to be placed (and resolved) within the context of the work setting and not some abstract model of the work process.

## 31.16 ETHNOGRAPHER AT WORK

The main rule is that methods that rely on retrospective accounts of social order cannot reveal members' methods. The method used must preserve the details of local order production "over its course" for the analyst. (Garfinkel 1967: 6)

the investigation of the rational properties of indexical expressions and other practical actions as contingent ongoing accomplishments of organised artful practices of everyday life. (Garfinkel 1967: 7)

Ethnomethodology's studies make vastly more sense when understood as inspections of the ways social scenes have visible coherence to even the most casual of witnesses, the ways in which the presence of social order can be readily detected within them; with the ways social order is exposed to even the most passing of glances... and reciprocally, the ways in which within such scenes the activities of individuals can be given definite sense, trajectory and motivation relative to the 'transparently' organised properties of the scene. (Sharrock and Button 1991: 163-164)

Ethnomethodology has consistently pointed to a yawning gap — the 'missing interactional what?' — in sociological studies of work that consists of all the missing descriptions of what occupational activities actually consist of and all the missing analyses of just how practitioners actually manage the workaday tasks which, for them, because they are workaday, are matters of mundane yet serious and pressing significance. For the ethnomethodologically informed ethnographer, there is no other place to stand in order to document, describe, and comprehend any setting than 'from the inside'. As Garfinkel argues:

"Ethnomethodologists generally use methods that require immersion in the situation being studied. They hold it as an ideal that they learn to be competent practitioners of whatever social phenomena they are studying." (Garfinkel 2002: 6)

Immersion in the milieu is a, if not '*the*', fundamental aspect of the ethnographer's work, and, in consequence, ethnographers spend considerable time developing 'unique adequacy' — learning to recognise and understand the activities and events that comprised the everyday world of work. In this fashion, the daily, mundane business of work, the conversations, asides, and acronyms become intelligible.

In getting to grips with and 'getting the hang of', the life of everyday work, researchers will necessarily learn various aspects of the practices and activities they are investigating — in some minimal sense actually how to do them — whatever 'them' happens to be, quilting, selling antiques, making bank loans, and so on. In that sense, ethnography presents the 'worm's eye' view of the world — since, generally, there are few conventionally 'important people' in everyday work. As P.J. O'Rourke (1989) reminds us, conventionally important people didn't get where they are by telling researchers the truth — nor should we ever fall for the sociologist's delusion — a variant of the 'Network Anchor-Creature self-conceit' that lets them, "believe Mikhail Gorbachev will suddenly take them aside and say, "Strictly between you and me, on Wednesday we invade Finland." (O'Rourke 1989:12. In some ways this is a necessary feature of ethnomethodologically informed ethnography, since arriving at an understanding of the social order from within requires documenting the 'worm's eye' view — producing thick descriptions of everyday activities, the materials used, the reasoning deployed, etc. — the 'shopwork' and 'shoptalk' (Garfinkel 2002). The ethnography (and perhaps the skill), then, consists in observing and describing how everyday work is achieved, how people observably and reportably act together to produce the objective, orderly, 'reality of social facts'. As Lemert suggests : " ethnomethodology imposes the obligation to study the utterly practical methods by which notoriously ordinary people compose the rational grounds of their social orderings." (Garfinkel 2002: xi). Whatever the arguments surrounding analytic approaches to the study of work, the primary challenge would appear to be to develop some vulgar competence in the field.

Ethnomethodologically informed ethnography requires looking at how people conduct their work in real settings, interested in exactly how work is socially organised in that setting. This means looking at the actual working division of labour as routinely and ordinarily manifested in the persons' meaningful orientation to their work, not work as some idealised conception — "the focus is on embodied, endogenous, witnessable practices." (Garfinkel 1967: 7). Despite some heroic conceptions of the ethnographer, derived largely from social anthropology, the work is fundamentally dull and boring — like work is for most people. The overwhelming emphasis of routine ethnographic work — describing the mundane features of everyday work — comes right up against the fact that work for most people has a generally dull if not unpleasant quality:

For most employees work has a generally unpleasant quality. If there is little Calvinist compulsion to work among propertyless factory workers and file clerks, there is also little Renaissance exuberance in the work of the insurance clerk, freight handler or department store saleslady ... Such joy as creative work may carry is more and more limited to a small minority. (Mills 1953: 219)

The thankless task of the ethnographer is simply to report in adequate detail how people go about doing what they construe as the things to be done. As such, ethnography is very much a practical activity; the fieldwork material — collected using a field notebook and a tape-recorder — is not dictated by strategic methodological considerations, but by the flow of activity within the setting. It simply involves recording what anyone is doing, moment by moment. Evidently, this does not demand any special or arcane skills for obtaining access and information — just everyday politeness — 'do you mind if I watch you work?'; 'what did you do then?', and so on. Despite concerns about contamination of data, Hawthorne effects etc. by and large, in this kind of setting, people have to get on with their work — and this is exactly what they tend to do. As Hughes et al. (2000) note, despite the apparent lack of method, the fieldworker cannot really fail, for even a few days of fieldwork is likely to produce an abundance if not an excess of material, of 'data'. The practical (and not to be underestimated) exercise, then, becomes one of gathering the accumulated materials and assembling them into a reasonable account of the work in the setting as a 'real world, real time' set of arrangements.

Like every other ethnographer 'immersed' in a setting, it will probably be your experience that your understanding of that setting, and what was going on within it, will develop gradually over the course of the fieldwork. Like everyone else, you probably have some vague notions of how 'work' gets done in your particular setting — how quilts are made, how aircrafts are controlled in the sky, how banking is carried on etc. — but equally there will be many things which you do not adequately comprehend. And so you will develop, over the course of the ethnography, a fuller, more informed sense of what the ways of work are and of the character of everyday work in a particular setting.

## 31.17 QUESTIONING THE 'METHOD': SOME PROBLEMS OF ETHNOGRAPHY

Ethnography is not a method without problems, many of which have been well documented (Randall et al. 1994) generally focusing on the standard concerns of 'getting in, staying in, getting out' as well as issues of access and 'gatekeeping', reliability, validity and generalisation, and so on. Ethnography is not, and, indeed, does not claim to be, a methodological panacea; though (perhaps fortunately) many of the critiques are directed at sociological, as opposed to ethnomethodological, variants of ethnography.

In practical terms, and historically, ethnography has generally been limited to small scale, well defined, and usually quite confined contexts, well suited to the observational techniques employed. Consequently, problems can arise with the method's application to large scale, highly distributed organisations. Similarly, in small scale settings there tends to be a clear focus of attention for the participants, who are typically few in number, and there is a relatively clearly visible differentiation of tasks at one work site. Scaling such inquiries up to the organisational level or to processes distributed in time and space is a much more daunting prospect.

In a similar vein, historically ethnography has been a 'prolonged activity', and whilst 'quick and dirty' approaches have been developed, the time scales involved in ethnographic research are often unrealistic in a commercial setting where the pressure is typically for 'results yesterday'. Moving out of the research setting into a more commercial one also raises different sets of ethical responsibilities as well as making access to sites more vulnerable to the contingencies of the commercial and industrial world. Ethnography insists that its inquiries should be conducted in a non-disruptive and non-interventionist manner — principles that can be compromised given that much of the motivation for introducing IT into the workplace is to reorganise work and, sometimes as part of this, to displace or deskill labour.

Since the 1970s, and particularly in recent years, the use of ethnography as a legitimate and viable research method has been challenged on various grounds — in particular that it privileges a white, western, male 'gaze'.

The questions were political, epistemological and methodological; who gets to say what about whom, and why? What are the interests and motivations behind *alleged* ethnographic 'realism? (Edles 2002: 145)

From within anthropology, ethnography has been accused of promoting a colonialist attitude (Said 1978) telling us more about the researchers, and their (usually his) attitudes, than the cultures they purport to describe. Within sociology, this kind of attack and charge — in this case of 'androcentricity' — has been endlessly repeated by various feminist writers (Reinharz 1992; Clough 1992), who suggest that ethnographies have mainly been conducted by males and are about males ignoring the role of women in the social setting. Clough (1992) for example suggests that an 'Oedipal logic' pervades traditional, realist ethnography, an ethnography that is effectively saturated with 'unconscious desire' — the desire to 'probe and penetrate' the world.

From within the ethnographic establishment, Hammersley (1990) has argued that the tendency to treat ethnographic description as involving simple reproduction of the phenomena described is misleading and mythical. He stresses that such description is always selective. Consequently, and following the 'reflexive turn', he suggests that the relevances and values that structure any ethnographic description must be made explicit. While it may be the case that ethnography retains an incoherent conception of its own goals and may frequently be a vehicle for ideology, such problems can be accepted without abandoning ethnography or its claims to represent phenomena — what he terms "subtle realism".

## 31.18 THE ETHNOGRAPHIC CRITIQUE OF ETHNOGRAPHY

While ethnography has always been subject to criticism from quantitative sociologists, as Brewer (1994) notes, it has recently come under attack from sociologists sympathetic to the method — the ethnographic critique of ethnography. This critique questions the reliability of ethnographic descriptions, and shows ethnographic texts to be artefacts, skilfully manufactured in order to construct their persuasive force. Continuing this line of argument, the postmodern critique of ethnography questions its claims to 'neutral realism', arguing that in writing ethnography, the researcher does not merely uncover or detail reality, but creates it in the interpretive process of creating the text, since 'reality' does not exist to be discovered. The 'textuality' debate has historical roots in philosophy and critical theory, but has recently culminated in the 'ethnographies as texts' movement and a lack of confidence in cultural description, what Marcus and Fischer (1986) refer to as a "crisis of representation" and Hammersley (1992) as a "crisis of fragmentation" in the ethnographic tradition. Clifford and Marcus, for example, argues that ethnographic writing is determined contextually, rhetorically, institutionally, generically, and historically, and that these "govern the inscription of coherent ethnographic fictions" (Clifford and Marcus, 1986: 6). In this view, the notion of a 'naturalist' ethnography that merely describes 'the facts of the matter' should instead be regarded as, "an insidious discursive strategy whose underlying purpose is to assert authority, dominate, and maintain privilege." (Edles 2002: 151) The reaction against 'naturalistic ethnography' — 'postmodern ethnography' — involves a mixture of literary styles, fiction, and poetry as part of faithfully representing the lived qualities of the domain. This response may also be seen as a reflexive device, collapsing the distinction between 'object' and 'subject' thereby facilitating ways of ensuring that authors write themselves into the text. This 'self-reflexive turn' takes a number of guises but often appears to take a confessional form whereby researchers document their own actions, attitudes, and prejudices and consider how this might have impacted on the setting they investigate.

This postmodern, constructivist challenge to naturalistic or 'naïve' ethnography and the subsequent demands for 'reflexive' ethnography, with a more self-critical and sceptical orientation has been challenged by those who conduct ethnomethodologically informed ethnographies (Sharrock 1995; Slack 2000).

The fact which has impacted upon both anthropologists and sociologists is that ethnography is, in important respects, perhaps even in essence … writing, and, as such is presumptively …exposed to deconstruction, to having its hidden agenda revealed, to its constituent texts being revealed to be self-defeating compositions.(Slack 2000)

Dicks et al (2005) suggest that recent writing on ethnography has focused on making it more attuned to reflecting complexity — in the form of contingency, multi-vocality, intertextuality, hybridity, and so on. They identify two aspects of 'post-paradigm' ethnographic enquiry in particular, the demarcation of ethnography's object of study and its mode of presentation, as areas of debate.

The category of ethnography, a well- established approach to social research in anthropology and some schools of sociology .. has been undergoing a continual process of diversification and fragmentation over the past 20 or so years. This has given rise to a variety of standpoints. It is now possible to identify an almost carnivalesque range of approaches under the ethnographic umbrella (Dicks et al 2005: 27)

In documenting the 'retreat of the author' and the development of a range of textual strategies, Dicks et al argue that ethnography is riddled with radical doubt.

Throughout these various standpoints runs a discursive turn, treating as central but problematic the relations of language, knowledge and power. Many of these perspectives indeed give rise to analyses that render ethnography itself — at least in any conventional mode — highly problematic, if not all-but-impossible.

(Dicks et al 2005: 27).

At the same time there has been the questioning of the category of 'the field', with its notion of easily identifiable spatial, geographical, and cultural boundaries.

## 31.19 CONSTRUCTIVIST CHALLENGES

As suggested earlier, postmodern constructivist challenges can be located within sociology's longstanding and notorious tradition of 'debunking', from which ethnomethodology fundamentally dissents. Constructivists seek to dispute the 'common sense' understandings that members of society have, often amounting to the bizarre suggestion that members of society really do not know what they are doing (and require a sociologist to tell them). The task constructivism sets itself is, of course, to challenge members' understandings, to show how they are wrong and present alternative, and authoritative, conceptions of both the way things are and how they got to be that way. These studies claim to show that what appears to members as common sense or obvious, for example that death or disability is a physical and biological event, is nothing of the kind but instead interpretative constructions, that can, therefore, be constructed differently, so that death or disability becomes a 'social construction'. (Grint and Woolgar 1992; Shakespeare 1993). Sharrock, following Bittner, views this development as part of the reaction against the concept of 'objectivity'.

The reaction against 'objectivity' ...was to move in a 'subjectivist' direction, to denounce all notions of objectivity, and to purport to root social phenomena in and to explore the dimensions of subjectivity. These tendencies were, in effect, to deny the existence of *social* reality, to make social reality a matter of *individual* determination - it was up to individuals to define social reality as they will.(Sharrock 1995: 13)

The result of this move, however, has been a shift away from a careful concern with the research setting and its members to a focus on the researcher and the research act itself — and the subsequent endless 'navel gazing', 'confessional tales', and piss-poor attempts at poetry (or jazz).

The constructivist view contrasts, then, with ethnomethodology's approach of indifference that attempts neither to undermine nor to support the everyday realities to which the members subscribe, but to investigate, describe, and understand them. As Sharrock argues:

Bittner, arguing on behalf of ethnomethodology, sought to distance it from just those tendencies, and to do so by arguing that the retreat from 'objectivity' as defined by those in the positivist traditions should not be toward 'subjectivity' but toward 'realism' - not realism, in the metaphysical sense, of asserting the existence of an external reality, but 'realism' in the phenomenological sense of faithfulness to the portrayal of its subject matter, a devotion to capturing society as it is actually experienced 'from within.' (Sharrock 1995: 15)

Bittner suggests that fieldwork strategies that have focused on detailing the experiences of the researcher are inclined to perpetuate this impoverishment in the portrayal of members' experience and represent a move away from a faithful description and rendering of the experience of members. At the same, time such ethnographies neglect the differences in the nature of the experiences of fieldworker and member. The supposition that 'social reality' is somehow grasped through the elaboration of the fieldworker's own awareness fundamentally misrepresents the very nature of the fieldworker's experience and motivation — as merely a 'visitor' that can return to a previous life. In this way, phenomena, the everyday occurrences in the setting, are divested of their massive sense of reality to those who routinely and necessarily inhabit that setting. Bittner's argument, that the ethnographic turn to 'subjectivity' involves increasing, almost exclusive, emphasis on the fieldworker's experience and point of view has been readily confirmed by the growing chorus for 'reflexivity' in sociology in general and ethnography in particular. (May and Perry 2010; Woolgar and Ashmore 1988). However, the emphasis on the fieldworker's standpoint as the focus for consideration of how social reality is engendered tends to overlook the extent to which the fieldworker's point of view is a peculiar one. While ethnographers may attempt to sensitise themselves to members' points of view, as Sharrock reminds us:

the fieldworker's occupation of that point of view is a temporary matter, .. The fieldworker does not, however, characteristically *occupy* the point of view ...The fieldworker *simulates* certain aspects of that view, but adopts it only for the purposes of the research, and as one which is freely taken up and from which it is equally possible readily to withdraw. (Sharrock 1995: 12)

In contrast, for members their 'native' point of view is not something to which they have a contingent relationship, one that they may freely take up, abandon, or exchange. In the setting of a bank, or air traffic control (or anywhere else) for example, the 'native point of view' is *their life,* something they have to take very seriously and not something they can 'play' with or relate to on a 'take it or leave it' basis. In a bank, the ways in which matters appear, for instance, to a bank manager — for example in terms of loans, overdrafts, repayments and so on — are mandatory for the manager and for others organisationally involved in the situation — these are the objective and (legally) binding ways of bank work. As a highly distributed organisation, the bank is reliant on the manager (and all its officials) acting in particular ways — indeed it can be a disciplinary matter if he fails to act accordingly. Bank personnel as a general rule cannot, except in their dreams (and often not then), playfully adopt a different point of view just to see what would happen, and the idea that things 'could be otherwise' is a possibility too childish for them to entertain. The playfulness of postmodernity rarely features as part of everyday work inside a bank, or most other commercial organisations.

As Gould et al. (1974) note, there are particular problems in ethnography's claim to describe events as they are seen or experienced by social actors. Asking people to explain what they are doing turns members into informants (Sacks 1992) and produces a 'perspective of action' (Gould et al. 1974) whereby settings are made meaningful to outsiders rather than a 'perspective in action' where meaning unfolds in naturally occurring interaction. Furthermore, there are some difficulties involved in seeking to understand the actor's perspective.

They treat as a 'perspective' what actors on most occasions view as the way the world is. The field worker, then, does not produce a description from the actor's point of view, but a description of the actor's point of view from the point of view of a sociological observer. This is true even if the observer seeks to empathise closely with actors' concerns and meanings. As a consequence, field-work descriptions tend to depict social life as perceived events and meanings, ignoring of distorting the lived reality of actor's worlds. (Emerson 1981: 357)

The emphasis in recent ethnographic writing on the 'reflexive' experience of the fieldworker, in that the fieldworker's history, attitudes, sexuality, etc. impacts on his or her perception of the setting leads to an under-estimate of the extent to which the experience of those under study possesses traits of depth and stability. In these circumstances, notions that 'it could have been — it could be — otherwise' are sociological fantasies. However, to critique constructionism is not a recommendation for accepting accounts at face value. Ethnomethodologically informed ethnographers choose instead to adopt a stance of 'indifference' to such questions, so issues of questioning or supporting an account do not arise. Thereby issues of truth and falsity and the endless debates of objectivity/subjectivity, the possibility of value neutrality, the researcher-researched relationship, and more are avoided. When considered from the viewpoint of sociological research, 'social reality' is clearly not the same thing as 'social reality for the purposes of everyday life'. As previously suggested, the actor cannot, under the auspices of the natural attitude, systematically adopt the sceptical stance found under the auspices of the theoretical attitude — we accept, rather than systematically doubt, everyday appearances.

However, this concern with the 'native's' point of view', with the difficulties of uncovering, displaying, and understanding a setting and way of life that is different, if not alien to the researcher, can also produce some unfortunate arguments about both how ethnographic research can be done and who is entitled to do it. The argument begins by suggesting, often quite rightly, that particular people's experience of research *on* them has often been less than happy. It is suggested that conventional ethnographic methods ignore the thoughts, feelings, and views of those they are researching — such as women, the disabled, ethnic minorities, and so on — thereby becoming one further aspect of disadvantage. (Dartington et al. 1981; Miller and Gwynne 1972)

Disabled people have come to see research as a violation of their experience, as irrelevant to their needs and as failing to improve their material circumstances and quality of life. (Oliver 1992: 105)

What is required, so the argument goes, are empathic research methods, deployed by those sympathetic to and experienced in the particular setting because, and here comes the extra twist, the researchers are themselves 'members' — disabled, women, people from an ethnic minority. So, the argument seems to shift from one about methods to one about who is warranted or entitled or qualified to conduct research. Again, it has to be acknowledged that this is hardly a unique argument but draws, for example, on long standing issues in feminist research and the critique of 'malestream' sociology. This includes disputes about not just what is investigated, but how research is conducted; arguments about 'objectivity', 'subjectivity' etc.; involvement of the 'subject ' in research; 'rape models', and so on.

Fortunately, ethnomethodologically informed ethnography avoids these debates by refusing to buy into many of the dichotomies of traditional social science — objective/subjective; structure/agency; etc. — that create many of these problems in the first place. In our view, the production of valid and useful ethnographic accounts relies initially on the satisfaction of the unique adequacy requirement. This insists that the researcher develops a vulgar competence in the setting itself in order to understand life as practitioners themselves comprehend and practice it and to be able to use the language of the setting to describe the setting. As Garfinkel and Weider put it:

for analysts to recognize, or identify, or follow the development of, or describe phenomena of order in local production of coherent detail the analyst must be vulgarly competent in the local production and reflexively natural accountability of the phenomena of order he [or she] is 'studying'. (Garfinkel and Weider 1992: 182)

As is sometimes argued, the issue is one of ‘probativeness’ (Garfinkel and Weider 1992) or of descriptive adequacy. In this case, at least understanding culture requires little more than a mundane competence in the practices of the domain such that the researcher can deliver an account that is intelligible to competent members. This is far from arguing that anyone who is not a bank worker (scientist, disabled, woman) is unable to write about, analyse, discuss, theorise, etc. these matters. ‘This crap’, as Jeff Coulter once said, ‘has got to stop’ (Crabtree 2000).

In ‘On the Demise of the Native’, Sharrock and Anderson (1982) point to some of the other problems of this kind of argument and approach. The argument behind the claim to exclusive access to a research setting confuses experience with understanding since it suggests that unless researchers possess the same ‘frameworks of meaning’ or experience, they cannot appreciate the everyday reality of members, and their research is correspondingly flawed. But this position — that particular members share a ‘culture’ that is different and inaccessible to others — is not only ludicrous but less a *finding* of research than an *a priori* principle. It is an assumption, not a discovery. Furthermore, the idea of a bank ‘culture’, of a shared set of meanings and understandings should be the endpoint of the analysis, i.e., the end product of serious and sustained enquiry, and not what enquiry is simplistically predicated upon. Essentially, the problem is posed as that of understanding an ‘alien’ culture. In this view, culture is all encompassing and people are regarded as empty vessels into which culture is somehow poured, and, in consequence, people end up both doing and knowing the same things. The ethnographer cannot understand this culture because s/he is not part of it. However, if we suspend this *a priori* status and make serious enquiries into that culture, we may well discover that what appear to be, or are represented as, massive cultural differences are, in fact, no more than variations in the ways some things are carried out. Understanding ‘bank culture’ or ‘football culture’ is not akin to the problem that Wittgenstein famously referred to when he stated, “if a lion could speak, we couldn't understand him" (Wittgenstein 1958: 223), but simply different ways of doing 'the same old thing'. For Wittgenstein, what we know and how we communicate is a function of our 'form of life' and thus understanding is embedded in our 'culture'. We understand because in our daily lives we live by routines — have cups of tea, use the computer, and so boringly on. Furthermore, if as Sharrock and Anderson (1982) suggest, the task of research is to demonstrate how culture and shared understanding is achieved, then the 'native' — in this case the worker — as well as the researcher should be regarded as enquirers into culture. In this circumstance, 'what is going on' becomes a problem for the native as well as the researcher, and the methods by which understanding is achieved are the focus of research.

.the stance that treats the native as an expert in his culture, knowing what he is up to and unproblematically recounting that to the researcher, may not be of much use. If we begin by positing that natives and researchers have to discover what is going on - what events and activities mean - then we can treat meaning as an achievable phenomenon and understanding as a risky business. It is these contingencies and risks that natives and field workers have to deal with. (Sharrock and Anderson 1982:135)

## 31.20 ETHNOGRAPHY AND REFLEXIVITY

Coffey (1999) argues that 'the self', and ethnographic subjectivity, as a pervasive feature of ethnographic enquiry have been ignored in the presentation of ethnography as an objective naturalistic form of research. Her focus is on the interaction between the researcher and the researched and how, "fieldwork shapes and constructs identities, intimate relations, an emotional self, and a physical self". Her argument, and it's an increasingly popular one, is that only by focusing on the researcher can the dualities that shape research and sociology be overcome. The ethnomethodological take on reflexivity is, not surprisingly, rather different. The fact that the term 'reflexivity' appears in Garfinkel's earlier formulations of ethnomethodology does not indicate *any* affinity between this use and its contemporary employment in talk of, for example, 'reflexive ethnography'. For ethnomethodologists, the notion of 'reflexivity' is best outlined in Garfinkel's classic description of accountability:

In exactly the ways that a setting is organised, it consists of members' methods for making evident that setting's ways as clear, coherent, planful, consistent, chosen, knowable, uniform, reproducible connections, - i.e., rational connections. In exactly the way that persons are members to organised affairs, they are engaged in serious and practical work of detecting, demonstrating, persuading through displays in the ordinary occasions of their interactions the appearances of consistent, clear, chosen, planful arrangements. In exactly the ways in which a setting is organised, it consists of methods whereby its members are provided with accounts of the setting as countable, storyable, proverbial, comparable, picturable, representable - i.e. accountable events. (Garfinkel 1967: 34)

For ethnomethodologists, fashionable concerns with 'reflexivity' are an irrelevance since our interest is fixated on production problems and the ways practices are produced and reproduced. The fixation is on visible orderliness, and our observations identify and describe 'grossly observable' phenomena — available to just about anyone. While for many sociologists the issue of 'reflexivity' is endlessly fascinating, inviting all kinds of what fundamentally amounts to 'navel gazing', for ethnomethodologists the reflexivity issue is entirely different since the emphasis is not on reflexivity of actors but reflexivity of accounts.

Slack (2000) argues that debates on reflexivity have "missed the need to ground their claims in the life world of society members". Slack makes the important distinction between what he terms 'essential' and 'stipulative' reflexivities. He suggests that stipulative reflexivity, "a sociological achievement",(Slack 2000: 1.2) has been the main concern of sociological researchers concerned to remedy members' versions of everyday life by attention to the analyst's perspective; "what counts as reflexivity is an achievement of the sociologist for sociology."

(ibid) Such an approach is based on a 'correspondence' epistemology — whereby 'reflexivity' permits or facilitates correct views of the social world. What such versions fail to recognise is that, in contrast the ethnomethodological approach to reflexivity — essential reflexivity — attends to members' reflexivity and is grounded in members' observable-reportable natural language practical actions. This emphasis, grounded in a 'coherence' epistemology, argues that there is no need for a sociological re-description, and that "the only way out of the postmodernist, structurational and textual maze is to attend to the practical essential reflexivity of society members." (ibid)

To briefly conclude this argument, the ethnomethodological endeavour lies in describing how *members* (not researchers or sociologists) manage to produce and recognise *contextually relevant* structures of social action. The warrant for ethnomethodologically informed ethnography is that of 'probativeness' or 'faithfulness to the phenomena' — that the description of the situated organisation of that activity *in its detail* makes that real worldly activity mutually intelligible.

## 31.21 IS ETHNOGRAPHY A 'METHOD' AT ALL?

Our answer to this is simple and unequivocal. No. This bears unpacking. Firstly, 'method' can be understood as entailing stepwise, logically related and ordered procedures, and ethnography clearly does not. It is not science, experiments are not conducted, variables are not controlled, and hypothesis testing (for the most part) is not done. More importantly, from our point of view, the emphasis on method is what has given rise to the immense and rather tedious literature we refer to above. Professional sociologists, let us not forget, have a vested interest in persuading others of their methodological expertise. Once the principles associated with ethnomethodology are grasped, however, all of the problems of 'reflexivity' and so on, simply disappear. If we accept that we inhabit a known-in-common world, in which basic principles of social interaction are recognised by (almost) everyone; where misunderstandings can be repaired; and where we can continue to interact even when we do not share ideological commitments, then neither sociologists nor anyone else have a privileged picture of ‘what the world is like’. In turn, this means that method, in and of itself, is really not that important. We aim to collect data in as reasonable a fashion as we can, using whatever material is to be found and — because we have no claims to methodological purity — are careful to limit our analytic claims about the world to what we have seen and can reasonably infer in much the way we describe above with our ‘precepts’.

## 31.22 MOVING THE METHOD ON: DEVELOPMENTS IN ‘ETHNOGRAPHIC’ APPROACHES

Our comments about ‘method’ can be construed as somewhat cynical, but they are not. In fact, they open the way — methodologically speaking — to any number of different analytic approaches. This includes, for brief mention, the fashionable themes of auto-ethnography, virtual ethnography, ‘postmodern’ ethnography, meta-ethnography, and multi-sited ethnography, as well as any number of developments in ‘method’ associated with qualitative work of this kind, including ‘living labs’, ‘cognitive walkthroughs’, online interviewing, textual analysis, etc. etc.

There is a strong sense in which ethnography has become both accepted and successful by the employment of ethnographers and anthropologists by companies like Microsoft, Nokia, Xerox, etc.

Video 31.6: Ethnography or observational research.

*Courtesy of Gerry Katz. Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms. View full screen or download*

Video 31.7: Ethnomethodology: Yahoo Research - Bob Moore.

*Courtesy of Bob Moore. Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms. View full screen or download (0)*

Video 31.8: Motorola research: assisted shopping.



At the same time, 'ethnographic' approaches in CSCW and HCI have begun to change as computing itself has changed to an interest in pervasive and ubiquitous computing, and the interest in technology has changed from a simple interest in productivity and profit to a range of more nebulous concerns such as fun and enjoyment and empathy and community, etc. The settings in which technology is deployed are increasingly sensitive and personal. Consequently, traditional, ethnographic, observational approaches have been supplemented by various forms of 'auto-ethnography' and devices such as 'technology probes' and 'cultural probes' (Gaver et al. 1996) and 'blogs' (Nardi et al. 2004; Graham et al. 2009).

Video 31.9: Cultural Probes - Qualitative Contextual Design Research.



Auto-ethnography has generated an enormous amount of comment, both approving and otherwise. Wrapped up in it are postmodern concerns with reflexivity and political objectives and objectification. Hence:

Autoethnography is . . . research, writing and method that connect the autobiographical and personal to the cultural and social. This form usually features concrete action, emotion, embodiment, self-consciousness, and introspection. (Ellis 2004: xix)

Autoethnography is . . . a self-narrative that critiques the situatedness of self and others in social context. (Spry 2001: 710)

Autoethnographic texts . . . democratize the representational sphere of culture by locating the particular experiences of individuals in tension with dominant expressions of discursive power. (Neumann 1996: 189)

Autoethnography is a blurred genre . . . a response to the call . . . it is setting a scene, telling a story, weaving intricate connections between life and art . . . making a text present . . . refusing categorization . . . believing that words matter and writing toward the moment when the point of creating autoethnographic texts is to change the world. (Jones 2005: 765)

It thus might entail personal narrative and experience; poetry, novelistic accounts, and politics. It addresses some obvious themes; that ethnography is never wholly 'innocent'; it can be used for 'standpoint' purposes (and has been, most notably in the context of disability studies); and it recognizes the essential reflexivity between ethnographer and his/her subject. But then, as Atkinson points out:

The list of ethnographic projects that draw on a personal commitment or accident is a long one and does need to be extended ad nauseam. There is, therefore, no need to rely exclusively on postmodernist rationales to justify such auto/biographical bases for ethnographic work. The ethnographer's identity and the subject matter of her or his chosen research site(s) have long been implicated in one another, and it is not a new development in the field of the social sciences.(Atkinson 2006: 401)

In passing, a nice (though journalistic) example of auto-ethnography of a kind is Rachel Simon's *Riding the Bus with my Sister* (2002).

One can say similar things about the other fashionable themes. There may well be some practical issues around how to study online behaviour (these are discussed *inter alia* by Hine 2000; Geiger and Ribes 2011), but they are not different *in kind*. Ethnographers have always had to contend with communication at a distance; with interrupted observation; with textual or documentary analysis, etc. The problems become unarguably more pronounced in certain circumstances, but they remain the problems of understanding interactional processes.

Much the same can be said of themes such as the 'multi-sited'. George Marcus's "Ethnography Through Thick and Thin" is the canonical text here. Marcus is an anthropologist who has been at the forefront of thinking about the nature of ethnography, the way in which ethnographic materials are presented or conveyed, and what 'usages' ethnography can be put to for some time. He asserts the view that ethnography needs to be understood as always being driven by particular analytic foci. In particular, he wants to challenge the 'realist' views of traditional anthropology (and the Chicago school sociology) by arguing that the future lies in interdisciplinarity. The point he is trying to make (we suggest) is that,

anthropologists have historically conducted their trade individually and this is one of the reasons for the decline in their authority.

interdisciplinarity will produce new analytic 'tropes' (themes or ideas).

Surprise, surprise, we broadly agree (thus far). The move to multi-sited ethnography, according to Marcus, is predicated on a number of factors, of which three seem particularly important:

Empirical changes in the world, notably in the global scope of capitalism and the technologies and artefacts that accompany it.

New forms of interdisciplinarity, and a concomitant crisis in the 'disciplines'.

The necessity for methodological responses that move beyond an apparent gap between the investigation of local detail and the theoretical concern with 'system' or 'structure'.

These general problems reflect significant changes in the modern/postmodern world. The development of new information and communication technologies which provide very rapid information flow; the rise of the 'global marketplace'; the globalisation of 'culture'; and the rise of new categories of homeless 'nomads' in which new 'structures of feeling', identities, or sensibilities become prevalent, have all in some way problematised the single site. Multi-sited ethnography, it is argued, might provide a response to these changes and theproblems they cause in a number of different ways, including prompting a new form of political and moral engagement, innovative methodological treatments, and a more sophisticated relationship between the construal of data and our understanding of the relevance of theory. Marcus outlines a so-called 'multi-sited' approach to ethnography. This represents, he thinks, a return to comparative ethnography, but in a different way:

comparison emerges from putting questions to an emergent object of study whose contours, sites and relationships are not known beforehand, but are themselves a contribution of making an account which has different, complexly connected real-world sites of investigation .... In the form of juxtapositions of phenomena that have conventionally appeared to be 'worlds apart. (Marcus 1998:86)

Marcus is careful to distinguish two modes through which, "... ethnographic research was embedding itself within the context of an historic and contemporary world system of capitalist political economy." Roughly, these modes correspond firstly to a procedure whereby single site ethnographic work which by other means contextualises the work in the world system. That is, data is collected locally to exemplify, 'fill in', or paint a portrait of a more global theoretical purpose. Less commonly, a 'postmodern' ethnography which moves out from single sites and which 'acknowledges macrotheoretical theories and narratives of the world system but does not rely on them for the contextual architecture framing a set of subjects' (Marcus 1998: 80), is developing. Methodologically, "this mobile ethnography takes unexpected trajectories in tracing a cultural formation across and within multiple sites of activity that destabilizes the distinction, say, between 'lifeworld' and 'system' ..." (Marcus 1998: 80).

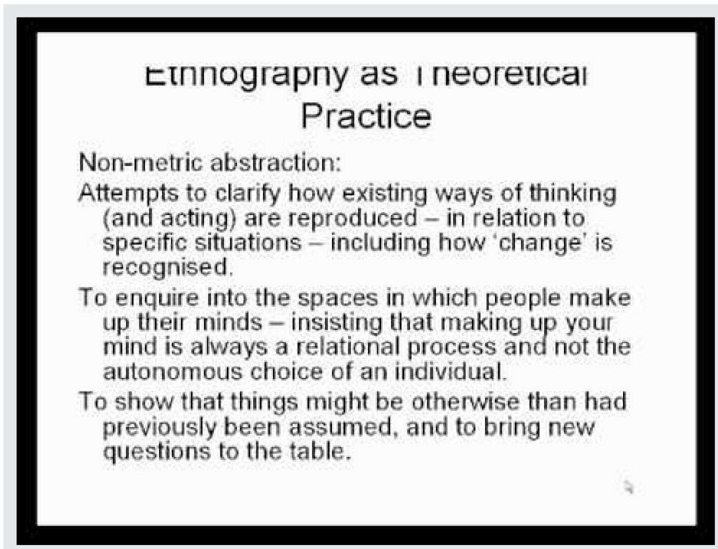
That is, multi-sited ethnography is in principle more than simply a methodological development. It plays with the dominant problematic of the social sciences. It is a problematic which is, indeed, as old as the social sciences and which has to do with the relation between structure and agency, and thus between data and theory — conceived in this instance as that of local and contextual detail, collected through some broadly 'qualitative' methods, and more or less 'grand' theoretical narratives. Moreover, a central feature of the problematic remains as it always was: the contest between 'scientific', 'positivist', and 'realist' modes of enquiry and the 'critical' form. In the latter, of course, theory is to be predominantly judged on critical value — the ability to critique and challenge the assumptions of 'normality' and the 'natural' that might otherwise be associated with, for instance, a global capitalist system.

The most important feature of this argument is that the problems of interdisciplinary engagement are not problems of method. Multi-sitedness implies an eclectic approach to 'method', and thus cannot (in any simplistic way) be about remedying the failure of other 'methods'. Nor are they problems of substantive disciplinary specific concerns because the contemporary crisis in those concerns is precisely what leads Marcus to interdisciplinarity. The 'multi-sited' view of interdisciplinarity, then, leads us to reflect on problems of *empirical* relevance, of *conceptual* orientation, and of the role of *comparison*.

Randall et al. have argued that these choices require "a particular open-mindedness about method, a thoughtful selection of concerns, and an artful refinement of disciplinary ... sensibilities." (Randall et al 2005: 82) Our point is tangential to this: debates about these matters are considerably less important than professional interests would have us believe. Since ethnomethodology is fundamentally anti-realist in its convictions, and shares something of Feyerabend's antipathy to method (see http://en.wikipedia.org/wiki/Paul_Feyerabend), all of these things can be fitted, though not all equally well, to its precepts. The standard is plausibility. When we tell a story about how people in some context organise their work activities, the information they rely on, and the things they are attentive to, we are not suggesting that we have dealt with all aspects of their world; that we have ordered them in importance; got everything right; couldn't have described things differently; or have been scrupulously 'objective'. What we are saying is that ethnography of the kind we practise is a simple thing, despite the protestations of professionals. We present our data in relation to our themes, open them for inspection so anyone can make observations about the degree to which they are useful, valid, truthful, or comprehensive, and say nothing at all about the things we haven't seen or cannot infer through ordinary common sense means. Let the sociologists talk among themselves.

## 31.23 SOME LINKS:

Manchester Methods




Video 31.10: What is ethnography? by Penny Harvey.

*Courtesy of Penny Harvey. Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms. View full screen or download*

Ethnography tutorial — the theory and practice of fieldwork and Guy Dewsbury's smartthinking site

## 31.24 REFERENCES


*(2013). Video: Motorola research: assisted shopping. Retrieved 13 June 2013 from http://www.interaction-design.org/tv/motorola_research_assisted_shopping.html*

*Agar, Michael H. (1980): The Professional Stranger: An Informal Introduction to Ethnography. New York, Academic Press*

*Anderson, R. J. (1994): Representations and Requirements: The Value of Ethnography in System Design. In Human-Computer Interaction, 9 (2) pp. 151-182*

*Anderson, Robert J., Hughes, John A. and Sharrock, Wes W. (1989): Working for Profit: The Social Organisation of Calculation in an Entrepreneurial Firm. Aldershot, United Kingdom,* *Avebury*

*Ashmore, Malcolm and Woolgar, Steve (1988): The next Step: An introduction to the reflexive project. In: Woolgar, Steve (ed.). "Knowledge and Reflexivity: New Frontiers in the Sociology of Knowledge". London, United Kingdom: Sage Publicationspp. 1-11*

*Atkinson, Paul (2006): Rescuing Autoethnography. In Journal of Contemporary Ethnography, 35 (4) pp. 400-404*

*Becker, Howard S. (1967): Whose Side Are We On?. In Social Problems, 14 (3) pp. 239-247*

*Becker, Howard S. (1953): Becoming a Marihuana User. In The American Journal of Sociology, 59 (3) pp. 235-242*

*Brewer, John D. (1994): The Ethnographic Critique of Ethnography: Sectarianism in the RUC. In Sociology, 28 (1) pp. 231-244*

*Button, Graham (2000): The Ethnographic Tradition and Design. In Design Studies, 21 (4) pp. 319-332*

*Button, Graham and Dourish, Paul (1996): Technomethodology: Paradoxes and Possibilities. In: Tauber, Michael J., Bellotti, Victoria, Jeffries, Robin, Mackinlay, Jock D. and Nielsen, Jakob (eds.) Proceedings of the ACM CHI 96 Human Factors in Computing Systems Conference April 14-18, 1996, Vancouver, Canada. pp. 19-26*

*Button, Graham and King, Val (1992): Hanging around is not the point. In: Proceedings of the 1992 ACM conference on Computer-supported cooperative work November 01 - 04, 1992, Toronto, Ontario, Canada.*

*Clifford, James and Marcus, George E. (eds.) (1986): Writing culture: the poetics and politics of ethnography. Los Angeles, USA, University of California Press*

*Clough, Patricia T. (1992): The End(s) of Ethnography: From Realism to Social Criticism. London, England, Sage Publications*

*Coffey, Amanda (1999): The Ethnographic Self: Fieldwork and the Representation of Identity. London, England, Sage Publications*

*Crabtree, Andy (2000): Methodological Issues Concerning the Practical Availability of Work-Practice to EM & CA. In: The 2nd Workplace Studies Conference October 27, 2000, Manchester, United Kingdom.*

*Crabtree, Andrew, Rodden, Tom, Tolmie, Peter and Button, Graham (2009): Ethnography considered harmful. In: Proceedings of ACM CHI 2009 Conference on Human Factors in Computing Systems 2009. pp. 879-888*

Dartington, Tim, Miller, Eric J. and Gwynne, Geraldine V. (1981): A Life Together: The Distribution of Attitudes around the Disabled. London, United Kingdom, Tavistock Publications

Dicks, Bella, Mason, Bruce, Coffey, Amanda and Atkinson, Paul (2005): Qualitative Research and Hypermedia: Ethnography for the Digital Age. London, United Kingdom, Sage Publications

Dingwall, Robert (1981): The Ethnomethodological Movement. In: Payne, Geoff, Dingwall, Robert, Payne, Judy and Carter, M. (eds.). "Sociology & Social Research". London, United Kingdom: Croom Helmpp. 124-138

Dourish, Paul (2013). Video: LIFT Conference || Paul Dourish (2008). Retrieved 13 June 2013 from http://www.interaction-design.org/tv/lift_conference_paul_dourish__2008.html

Dourish, Paul (2013a): Reading and Interpreting Ethnography. In: Kellogg, Wendy and Olson, Judy (eds.). "HCI Ways of Knowing".

Dourish, Paul (2006): Implications for design. In: Proceedings of ACM CHI 2006 Conference on Human Factors in Computing Systems 2006. pp. 541-550

Dourish, Paul and Button, Graham (1998): On "Technomethodology": Foundational Relationships between Ethnomethodology and System Design. In Human-Computer Interaction, 13 (4) pp. 395-432

Edles, Laura D. (2002): Cultural Sociology in Practice. Oxford, United Kingdom, Wiley-Blackwell

Ellis, Carolyn (2004): The Ethnographic I: A Methodological Novel about Autoethnography. Walnut Creek, USA, AltaMira Press

Emerson, Robert M. (1981): Observational Field Work. In Annual Review of Sociology, 7 pp. 351-378

Freeman, Derek (1999): The Fateful Hoaxing Of Margaret Mead: A Historical Analysis Of Her Samoan Research. Boulder,USA, Westview Press

Garfinkel, Harold (2002): Ethnomethodology's Program: Working Out Durkheim's Aphorism. Lanham, USA, Rowman and Littlefield Publishers

Garfinkel, Harold (1967): Studies in Ethnomethodology. London, United Kingdom, Polity Press

Garfinkel, Harold (1996): Studies in Ethnomethodology. London, United Kingdom, Polity Press

*Garfinkel, Harold and Weider, Lawrence (1992): Evidence for locally produced, naturally accountable phenomena of order, logic, reason, meaning, method, etc. In: Watson, Graham and Seiler, Robert M. (eds.). "Text in Context: Contributions to ethnomethodology". London, United Kingdom: Sage Publicationspp. 175-206*

*Gaver, Bill (2013). Video: Cultural Probes - Qualitative Contextual Design Research. Retrieved 13 June 2013 from http://www.interaction-design.org/tv/cultural_probes_-_qualitative_contextual_design_re.html*

*Gaver, William, Dunne, Tony and Pacenti, Elena (1999): Design: Cultural probes. In Interactions, 6 (1) pp. 21-29*

*Geiger, Stuart R. and Ribes, David (2011): Trace Ethnography: Following Coordination through Documentary Practices. In: 44th Hawaii International Conference On System Sciences January 4-7, 2011, Hawaii, USA. pp. 1-10*

*Goguen, Joseph A. (1993): Social issues in requirements engineering. In: International Symposium on Requirements Engineering January 4-6, 1993, San Diego, USA. pp. 194-195*

*Gold, Raymond L. (1958): Roles in sociological field investigation. In Social Forces, 38 (3) pp. 217-223*

*Gould, Leroy, L.Walker, Andrew, E.Crane, Lansing and W.Lids, Charles (1974): Connections: Notes From The Heroin World. New Heaven, USA, Yale University Press*

*Graham, Connor, Rouncefield, Mark and Satchell, Christine (2009): Blogging as 'therapy'? Exploring personal technologies for smoking cessation. In Health Informatics Journal, 15 (4) pp. 267-281*

*Grint, Keith and Woolgar, Steve (1992): Computers, guns, and roses: what's social about being shot. In Science, Technology, & Human Values, 17 (3) pp. 336-380*

*Gubrium, Jaber F. and Holstein, James A. (1997): The new language of qualitative method. New York, USA, Oxford University Press*

*Hammersley, Martyn (1990b): What's Wrong with Ethnography? The Myth of Theoretical Description. In Sociology, 24 (4) pp. 597-615*

*Hammersley, Martyn (1992): On Feminist Methodology. In Sociology, 26 (2) pp. 187-206*

*Hammersley, Martyn (1990a): The Dilemma of Qualitative Method: Herbert Blumer and the Chicago Tradition. London, United Kingdom, Routledge*

*Harper, Richard H., W.Randall, Dave, Smythe, N., Evans, C., Heledd, L. and Moore, R. (2007): Thanks for the memory. In: Ormerod, Tom and Sas, Corina (eds.) Proceedings of the*

*21st British HCI Group Annual Conference on People and Computers* September 3-7, 2007, Lancaster, United Kingdom. pp. 39-42

Harper, R., Randall, D., Smyth, N., Evans, C., Heledd, L. and Moore, R. (2008): The past is a different place: they do things differently there. In: *Proceedings of DIS08 Designing Interactive Systems* 2008. pp. 271-280

Harvey, Penny (2013). Video: What is ethnography? by Penny Harvey. Retrieved 13 June 2013 from http://www.interaction-design.org/tv/what_is_ethnography_by_penny_harvey.html

Heath, Christian and Luff, Paul (1992): Collaboration and control: Crisis management and multimedia technology in London Underground Line Control Rooms. In *Computer Supported Cooperative Work*, 1 (1) pp. 69-94

Hine, Christine (2000): Virtual Ethnography. London, United Kingdom, Sage Publications

Hughes, John and King, Val (): Sociology for Large-Scale Systems Design. In: *Proceedings of CRICT conference on Software and Systems* .

Hughes, John, O'Brien, Jon, Rodden, Tom and Rouncefield, Mark (2000): Ethnography, Communication and Support for Design. In: Luff, Paul, Hindmarsh, Jon and Heath, Christian (eds.). "Workplace Studies: Recovering Work Practice and Informing System Design". Cambridge University Presspp. 187-214

Hughes, John, King, Val, Rodden, Tom and Andersen, Hans (1994): Moving out of the Control Room: Ethnography in System Design. In: *Proceedings of the 1994 ACM conference on Computer supported cooperative work* October 22 - 26, 1994, Chapel Hill, North Carolina, United States. pp. 429-439

Hughes, John A., Randall, David and Shapiro, Dan (1992): Faltering from Ethnography to Design. In: *Proceedings of the 1992 ACM conference on Computer-supported cooperative work* November 01 - 04, 1992, Toronto, Ontario, Canada. pp. 115-122

Hughes, John A., O'Brien, Jon, Rodden, Tom, Rouncefield, Mark and Sommerville, Ian (1995): Presenting ethnography in the requirements process. In: *Second IEEE International Symposium on Requirements Engineering 1995* March 27 - 29, 1995, York, England. pp. 27-39

Hughes, John A., O'Brien, Jon, Rodden, Tom and Rouncefield, Mark (1997): Designing with ethnography: a presentation framework for design. In: McClelland, Ian, Olson, Gary, Veer, Gerrit van der, Henderson, Austin and Coles, Susan (eds.) *Proceedings of the 2nd conference on Designing interactive systems processes, practices, methods, and techniques* August 18-20, 1997, Amsterdam, Netherlands. pp. 147-158

*Jones, Stacy H. (2005): Autoethnography: Making the personal political. In: Denzin, Norman K. and Lincoln, Yvonna S. (eds.). "Handbook of Qualitative Research". Thousand Oaks, California, USA: Sage Publicationspp. 763-792*

*Katz, Gerry (2013). Video: Ethnography or observational research. Retrieved 13 June 2013 from http://www.interaction-design.org/tv/ethnography_or_observational_research.html*

*Lynch, Mike (2013). Video: Ethnomethodology: Mike Lynch on Ethnomethodological studies of work in the sciences. Retrieved 13 June 2013 from http://www.interaction-design.org/tv/ethnomethodology_mike_lynch_on_ethnomethodologica.html*

*Marcus, George E. (1998): Ethnography through Thick and Thin. Princeton, USA, Princeton University Press*

*Marcus, George E. and Fischer, Michael M. (1986): Anthropology as cultural critique. Chicago, USA, University of Chicago Press*

*Martin, D., Rodden, Tom, Rouncefield, Mark, Sommerville, Ian and Viller, S. (2001): Finding patterns in the fieldwork. In: Ecscw 2001 - Proceedings of the Seventh European Conference on Computer Supported Cooperative Work 16-20 September, 2001, Bonn, Germany. pp. 39-58*

*Martin, David, Rouncefield, Mark and Sommerville, Ian (2002): Applying patterns of cooperative interaction to work (re)design: e-government and planning. In: Terveen, Loren (ed.) Proceedings of the ACM CHI 2002 Conference on Human Factors in Computing Systems Conference April 20-25, 2002, Minneapolis, Minnesota. pp. 235-242*

*May, Tim and Perry, Beth (2010): Sociological Research and Reflexivity: Content, consequence and context. London, United Kingdom, Sage Publications*

*Miller, Eric J. and Gwynne, G.V. (1972): A Life Apart: Pilot Study of Residential Institutions for the Physically Handicapped and the Young Chronic Sick (Soc. Sci. Pbs.). London, United Kingdom, Tavistock Publications*

*Mills, Wright C. (1953): White Collar Work: The American Middles Classes. New York, USA, Oxford University Press*

*Moore, Bob (2013). Video: Ethnomethodology: Yahoo Research - Bob Moore. Retrieved 13 June 2013 from http://www.interaction-design.org/tv/ethnomethodology_yahoo_research_-_bob_moore.html*

*Nardi, Bonnie A., Schiano, Diane J. and Gumbrecht, Michelle (2004): Blogging as social activity, or, would you let 900 million people read your diary?. In: Proceedings of ACM CSCW04 Conference on Computer-Supported Cooperative Work 2004. pp. 222-231*

*Neumann, Mark (1996): Collecting ourselves at the end of the century. In: Ellis, Carolyn and Bochner, Arthur P. (eds.). "Composing ethnography: Alternative forms of qualitative writing". Walnut Creek,California, USA: AltaMira Presspp. 172-198*

*O'Rourke, P.J. (1989): Holidays in Hell. New York, USA, Vintage Books*

*Plowman, Lydia, Rogers, Yvonne and Ramage, Magnus (1995): What Are Workplace Studies For?. In: Marmolin, Hans, Sundblad, Yngve and Schmidt, Kjeld (eds.) ECSCW 95 - Proceedings of the Fourth European Conference on Computer-Supported Cooperative Work 11-15 September, 1995, Stockholm, Sweden. pp. 309-324*

*Pollner, Melvin and Emerson, Robert M. (2001): Ethnomethodology and Ethnography. In: Atkinson, Paul, Coffey, Amanda, Delamont, Sara, Lofland, John and Lofland, Lyn (eds.). "Handbook of Ethnography". London, United Kingdom: Sage Publicationspp. 118-136*

*Randall, Dave, Harper, Richard and Rouncefield, Mark (2005): Fieldwork and Ethnography: A Perspective From CSCW. In: Proceedings of the 1st International Ethnographic Praxis in Industry and Commerce Epic Conference 2005, Redmond, USA. pp. 81-99*

*Randall, Dave, Hughes, John and Shapiro, Dan (1994): Steps toward a partnership: ethnography and system design. In: Jirotka, Marina, Goguen, Joseph A., Monk, Andrew F. and Gaines, Brian R. (eds.). "Requirements Engineering". San Diego, USA: Academic Presspp. 241-258*

*Randall, David, Harper, Richard and Rouncefield, Mark (2007): Fieldwork for Design. Berlin,Germany, Springer*

*Reinharz, Shulamit and Davidman, Lynn (1992): Feminist methods in social research. New York, USA, Oxford University Press*

*Rittel, Horst W. and Webber, Melvin M. (1973): Dilemmas in a general theory of planning. In Policy Sciences, 4 (2) pp. 155-169*

*Sacks, Harvey (1992): Lectures in Conversation. Oxford, United Kingdom, Wiley-Blackwell*

*Shakespeare, Tom W. (1993): Disabled people's self-organisation: a new social movement?. In Disability & Society, 8 (3) pp. 249-264*

*Shankman, Paul (2000): Culture, Biology and Evolution: The Mead- Freeman controversy. In Journal of Youth and Adolescence, 29 (5) pp. 539-556*

*Shapiro, Dan (1994): The Limits of Ethnography: Combining Social Sciences for CSCW. In: Proceedings of the 1994 ACM conference on Computer supported cooperative work October 22 - 26, 1994, Chapel Hill, North Carolina, United States. pp. 417-428*

*Sharrock, Wes (2013). Video: Ethnomethodology by Wes Sharrock. Retrieved 13 June 2013 from http://www.interaction-design.org/tv/ethnomethodology_by_wes_sharrock.html*

*Sharrock, Wes W. and Anderson, Robert J. (1982): The demise of the native. In Human Studies, 5 (2) pp. 119-135*

*Sharrock, Wes W. and Anderson, Robert J. (1991): Epistemology: professional scepticism. In: Button, Graham (ed.). "Ethnomethodology and the Human Sciences". London, United Kingdom: Routledgepp. 51-76*

*Sharrock, Wes W. and Button, Graham (1991): The social actor: social action in real time. In: Button, Graham (ed.). "Ethnomethodology and the Human Sciences". London, United Kingdom: Routledgepp. 135-175*

*Sharrock, Wes W. and Hughes, John A. (2002): Ethnography in the Workplace: Remarks on its theoretical bases. In TeamEthno-Online, (1)*

*Sharrock, W. 1995 Issues in Ethnography: Ethnomethodology and Constructionism, COM-IC document, MAN-2-4*

*Silverman, David (1985): Qualitative Methodology and Sociology: Describing the Social World. Aldershot, United Kingdom, Gower Pub Co*

*Silverman, David (1993): Interpreting Qualitative Data: Methods for Analysing Talk, Text and Interaction. London, United Kingdom, Sage Publications*

*Simon, Rachel (2002): Riding the Bus With My Sister: A True Life Journey. New York, USA, Houghton Mifflin*

*Slack, Roger (2000). Reflexivity of Sociological Practice A Reply to May. Retrieved 10 April 2013 from Sociological Research Online: http://www.socresonline.org.uk/5/1/slack.html*

*Smith, Dorothy (2013). Video: Institutional Ethnography. Retrieved 13 June 2013 from http://www.interaction-design.org/tv/institutional_ethnography.html*

*Sommerville, Ian, Rodden, Tom, Sawyer, Pete, Bentley, Richard and Twidale, Michael (1993): Integrating ethnography into the requirements engineering process. In: IEEE International Symposium on Requirements Engineering RE93 January 6, 1993, San Diego, USA. pp. 165-173*

*Sommerville, Ian, Rodden, Tom, Sawyer, Pete and Bentley, Richard (1992): Sociologists Can be Surprisingly Useful in Interactive Systems Design. In: Monk, Andrew, Diaper, Dan and Harrison, Michael D. (eds.) Proceedings of the Seventh Conference of the British Computer Society Human Computer Interaction Specialist Group - People and Computers VII August 15-18, 1992, University of York, UK. pp. 341-353*

*Spry, Tami (2001): Performing autoethnography: An embodied methodological praxis. In Qualitative Inquiry, 7 (6) pp. 706-732*

Strathern, Marilyn (2003): Commons and Borderlands: Working Papers on Interdisciplinarity, Accountability and the Flow of Knowledge. Hereford, Sean Kingston Publishing

Suchman, Lucy A. (1983): Office Procedures as a practical activity: Models of work and system design. In Transactions on office Automation Systems, 1 (4) pp. 320-328

Suchman, Lucy A. (1995): Making Work Visible. In Communications of the ACM, 38 (9) pp. 56-64

Viller, Stephen and Sommerville, Ian (1999): Social Analysis in the Requirements Engineering Process: From Ethnography to Method. In: 4th IEEE International Symposium on Requirements Engineering RE 99 7-11 June, 1999, Limerick, Ireland. pp. 6-13

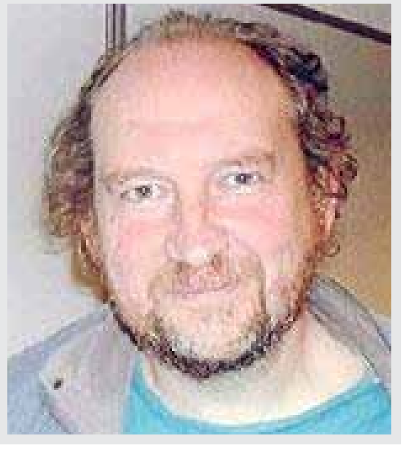
Williamson, Bill (1989): Review Article: Sentiment and Social Change. In The Sociological Review, 37 (1) pp. 128-141

Wittgenstein, Ludwig (1958): Philosophical Investigations. Oxford, United Kingdom, Blackwell

Wittgenstein, Ludwig (1984): Culture and Value. Chicago, USA, University of Chicago Press

## About the authors

### Dave Randall



Current place of employment:
University of Siegen, Germany (visiting)

Dave Randall was Principal Lecturer at Manchester Metropolitan University, UK. until his retirement in 2011. He sorely misses his administrative load. He continues to work, collaborating with people like his co-author on this piece; with Richard Harper at Microsoft Research where they are currently involved in writing a book on Choice, and with Volker Wulf at the University of Siegen in Germany. His work sits primarily in the interdisciplinary research area called Computer Supported Cooperative Work (CSCW)and in HCI. He is particularly interested in the application of the ethnomethodological 'studies of work' programme to problems of new technology and organizational change, and in the conduct of ethnographic enquiry in relation to these issues. He has conducted a number of studies of 'work in organizations' in his career. These include a well-known and extensively-cited study of Air Traffic Control as well as studies of retail financial services, museum work, classroom interaction with new technology, ontology-based design, mobile phone use, and 'smart home' technology. He has undertaken consultancy and other work with organizations such as the Riso national laboratory, Denmark; Xerox plc; the Children's Society; Orange plc; Vodaphone plc; Microsoft plc and the national Centre for E-Social Science (NCess) and has collaborated with partners in a number of other institutions in the UK and Europe over a period of time. These include Lancaster University; Manchester University; the Blekinge Institue of technology and Lulea Technical University in Sweden, and the I.T. University of Denmark. He has co-authored three books, one an examination of organizational change and new technology in the retail financial services sector and another (with Mark Rouncefield and Richard Harper) on the conduct of ethnography for design-related purposes. A third is also, oddly enough, co-authored with Mark Rouncefield (and others): 'Technologies of Leadership in F.E.'. He also has one edited book with two more on the way.Currently (2013), he is working on a report for the Economic and Social Research Council on virtual learning environments and with Hitachi Ltd in japan on the applicability of 'patterns' to engineering maintenance work.

## Mark Rouncefield

© *Mark Rouncefield*

Has also published under the name of:
"M. Rouncefield"

Personal Homepage:
[comp.lancs.ac.uk/~rouncefi/](comp.lancs.ac.uk/~rouncefi/)

Mark Rouncefield is a Senior Research Fellow in the Computing Department, Lancaster University, concerned with carrying out a number of ethnomethodologically informed ethnographic studies of Computer Supported Co-operative Work.

## Your notes and thoughts on chapter 31

Record your notes and thoughts on this chapter. If you want to share these thoughts with others online, go to the bottom of the page at:
http://www.interaction-design.org/encyclopedia/ethnography.html

**NOTES:**

# Chapter 32

# 32. 3D User Interfaces

***by Doug A. Bowman.***

Ever since the advent of the computer mouse and the graphical user interface (GUI) based on the Windows, Icons, Menus, and Pointer (WIMP) paradigm, people have asked what the next paradigm shift in user interfaces will be (van Dam, 1997; Rekimoto, 1998). Mouse-based GUIs have proven remarkably flexible, robust, and general, but we are finally seeing a major sea change towards "natural" user interfaces (NUIs), not only in the research lab, but also in commercial products aimed at broad consumer audiences. Under the NUI umbrella, there are two broad categories of interfaces: those based on direct touch, such as multi-touch tablets (Wigdor & Wixon, 2011), and those based on three-dimensional spatial input (Bowman et al., 2005), such as motion-based games. It is this latter category, which we call three-dimensional user interfaces (3D UIs), that we focus on in this chapter.

## 32.1 WHAT ARE 3D USER INTERFACES?

Like many high-level descriptive terms in our field (such as "virtual reality" and "multimedia"), it's surprisingly difficult to give a precise definition of the term "3D user interface." Although most practitioners and researchers would say, "I know one when I see one," stating exactly what constitutes a 3D UI and which interfaces should be included and excluded is tricky.

*3D User Interfaces: Theory and Practice* (Bowman et al., 2005) defines a 3D user interface as simply "a UI that involves 3D interaction." This simply delays the inevitable, as we now have to define 3D interaction. The book states that 3D interaction is "human-computer interaction in which the user's tasks are performed directly in a 3D spatial context."

One key word in this definition is "directly." There are some interactive computer systems that display a virtual 3D space, but the user only interacts indirectly with this space—e.g., by manipulating 2D widgets, entering coordinates, or choosing items from a menu. These are not 3D UIs.

The other key idea is that of a "3D spatial context." The book goes on to make it clear that this spatial context can be either physical or virtual, or both. The most prominent types of 3D UIs involve a physical 3D spatial context, used for input. The user provides input to the system by making movements in physical 3D space or manipulating tools, sensors, or devices in 3D space, without regard for what this input is used to do or control. Of course, all input/interaction is in some sense in a physical 3D spatial context (a mouse and keyboard exists in 3D physical space), but the intent here is that the user is giving spatial input that involves 3D position (x, y, z) and/or orientation (yaw, pitch, roll) and that this spatial input is meaningful to the system.

Thus, the key technological enabler of 3D UIs of this sort is spatial *tracking* (Meyer et al., 1992; Welch & Foxlin, 2002). The system must be able to track the user's position, orientation, and/or motion to enable this input to be used for 3D interaction. For example, the Microsoft Kinect tracks the 3D positions of multiple body parts to enable 3D UIs, while the Apple iPhone tracks its own 3D orientation, allowing 3D interaction. There are many different technologies used for spatial tracking; we describe some of these in a later section.

This tracked spatial input can be used for iconic gestures, direct pointing at menu items, controlling characters in a game, specifying 3D shapes, and many other uses. 3D UIs based on spatial input can be found in a variety of settings: gaming systems, modeling applications, virtual and augmented reality systems, large screen visualization setups, and art installations, just to name a few.

The other type of 3D UI involves direct interaction in a virtual 3D spatial context. In this type, the user may be using traditional (non-3D) input devices or movements as inputs, but if those inputs are transformed directly into actions in the virtual 3D space, we still consider it to be 3D interaction. For example, the user might drag the mouse across a 3D model in order to paint it a certain color, or the user might draw a path through a 3D world using touch input.

In this , we are going to focus on the first type of 3D UI, which is based on 3D spatial input. While both types are important and have many applications, they involve different research issues and different technologies to a large degree. 3D spatial tracking has come of age recently, and based on this technological driver, 3D UI applications with spatial input have exploded. We discuss a few of these applications in more detail in the next section.

## 32.2 APPLICATIONS OF 3D UIS

Why is it important to understand and study 3D UIs? For many years, the primary application of 3D UIs was in high-end virtual reality (VR) and augmented reality (AR) systems. Since users in these systems were generally standing up, walking around, and limited in their view of the real world, traditional mouse- and keyboard-based interaction was impractical. Such systems were already using spatial tracking of the user's headthe correct view of the virtual world, it was natural to also design UIs that took advantage of spatial tracking as well. As we indicated above, however, recent years have seen an explosion of spatial input in consumer-level systems such as game consoles and smartphones. Thus, the principles of good 3D UIs design are now more important to understand than ever.

To further motivate the importance of 3D UI research, let's look in a bit more detail at some important technology areas where 3D UIs are making an impact on real-world applications.

### 32.2.1 Video Gaming

As we've already mentioned, most people today are aware of 3D UIs because of the great success of "motion gaming" systems like the Nintendo Wii, the Microsoft Kinect, and the Sony Move. All of these systems use spatial tracking to allow users to interact with games through pointing, gestures, and most importantly, natural movements, rather than with buttons and joysticks. For example, in an archery game a user can hold two tracked devices—one for the handle of the bow and the other for the arrow and string—and can pull back the arrow, aim, and release using motions very similar to archery in the real world.

The Wii and Move both use tracked handheld devices that also provide buttons and joysticks, while the Kinect tracks the user's body directly. There's a clear tradeoff here. Buttons and joysticks are still useful for discrete actions like confirming a selection, firing a weapon, or changing the view. On the other hand, removing encumbrances from the user can make the experience seem even more natural.

3D UIs are a great fit for video gaming (LaViola, 2008; Wingrave et al., 2010), because the emphasis is on a compelling experience, which can be enhanced with natural actions that make the player feel as if he is part of the action, rather than just indirectly controlling the actions of a remote character.

### 32.2.2 Very Large Displays

Recent years have seen an explosion in the size, resolution, and ubiquity of displays. So-called "display walls" are found in shopping malls, conference rooms, and even people's homes. Many of these displays are passive, simply presenting canned information to viewers, but more and more of them are interactive.

So how should one interact with these large displays? The traditional mouse and keyboard still work, but they are difficult to use in this context because users want to move about in front of the display, and because such large displays invite multiple users (Ball and North, 2005). Touch screens are another option, but that means that to interact with the display one has to stand within arm's reach, limiting the amount of the display that can be seen.

3D interaction is a natural choice for large display contexts. A tracked handheld device, the hand itself, or the whole body can be used as portable input that works from any location and makes sense for multiple users. The simplest example is *distal pointing*, where the user points directly at a location on the display (as with a laser pointer) to interact with it (Vogel & Balakrishnan, 2005; Kopper et al., 2010), but other techniques such as full-body gestures or viewpoint-dependent display can also be used.

### 32.2.3 Mobile Applications

Today's mobile devices, such as smartphones and tablets, are an interaction designer's playground, not only because of the rich design space for multi-touch input, but also because these devices incorporate some fairly powerful sensors for 3D spatial input. The combination of accelerometers, gyroscopes, and a compass give these devices the ability to track their own orientation quite accurately. Position information based on GPS and accelerometers is less accurate, but still present. These devices offer a key opportunity for 3D interaction design, however, because they are ubiquitous, they have their own display, and they can do spatial input without the need for any external tracking infrastructure (cameras, base stations, etc.).

Many mobile games are using these capabilities. Driving games, for example, use the “tilt to steer” metaphor. Music games can sense when the user is playing a virtual drum. And golf games can incorporate a player’s real swing.

But “serious” applications can take advantage of 3D input for mobile devices as well. Everyone is familiar with the idea of tilting the device to change the interface from portrait to landscape mode, but this is only the tip of the iceberg. A tool for amateur astronomers can use GPS and orientation information to help the user identify stars and planets they point the device towards. Camera applications can not only record the location at which a photo was taken, but also track the movement of the camera to aid in the reconstruction of a 3D scene.

Perhaps the most prominent example of mobile device 3D interaction is in mobile AR. In mobile AR, the smartphone becomes a window through which the user can see not only the real world, but virtual objects and information as well (Höllerer et al., 1999; Ashley, 2008). Thus, the user can browse information simply by moving the device to view a different part of the real world scene. Mobile AR is being used for applications in entertainment, navigation, social networking, tourism, and many more domains. Students can learn about the history of an area; friends can find restaurants surrounding them and link to reviews; and tourists can follow a virtual path to the nearest subway station. Prominent projects like MIT’s SixthSense (Mistry & Maes, 2009) and Google’s Project Glass (Google, 2012) have made mobile AR highly visible. Good 3D UI design is critical to realizing these visions.

## 32.3 3D UI TECHNOLOGIES

As we discussed above, spatial tracking technologies are intimately connected to 3D UIs. In order to design usable 3D UIs, then, a basic understanding of spatial tracking is necessary. In addition, other input technologies and display devices play a major role in 3D UI design.

### 32.3.1 Tracking Systems and Sensors

Spatial tracking systems sense the position, orientation, linear or angular velocity, and/or linear or angular acceleration of one or more objects. Traditionally, 3D UIs have been based on six-degree-of-freedom (6-DOF) position trackers, which detect the absolute 3D position (location in a fixed XYZ coordinate system) and orientation (roll, pitch, and yaw in the fixed coordinate system) of the object, which is typically mounted on the head or held in the hand.

These 6-DOF position trackers can be based on many different technologies, such as those using electromagnetic fields (e.g., Polhemus Liberty), optical tracking (e.g., NaturalPoint OptiTrack), or hybrid ultrasonic/inertial tracking (e.g., Intersense IS900). All of these, however, share the limitation that some external fixed reference, such as a base station, a camera array, a set of visible markers, or an emitter grid, must be used. Because of this, absolute 6-DOF position tracking can typically only be done in prepared spaces.

Inertial tracking systems, on the other hand, can be self-contained and require no external reference. They use technologies such as accelerometers, gyros, magnetometers (compasses), or video cameras to sense their own motion—their change in position or orientation. Because they measure relative position and orientation, inertial systems can't tell you their absolute location, and errors in the measurements tend to accumulate over time, producing drift.

The "holy grail" of spatial tracking is a self-contained 6-DOF system that can track its own absolute position and orientation with high levels of accuracy and precision. We are getting closer to this vision. For instance, a smartphone can use its accelerometers, gyros, and magnetometer to track its absolute orientation (relative to gravity and the earth's magnetic field), and its GPS receiver to track its 2D position on the surface of the earth. However, GPS position is only accurate to within a few feet at best, and the height (altitude) of the phone cannot currently be tracked with any accuracy. For now, then, smartphones on their own cannot be used as a general-purpose 6-DOF input device.

A 6-DOF tracker with minimal setup requirements is the Sony Move system. Designed as a "motion controller" (although it really senses position) for the PlayStation game console, the Move uses the typical accelerometers and gyros to sense 3D orientation, and a single camera to track the 3D position of a glowing ball atop the device. This works surprisingly well, coming near to the accuracy of much more expensive and complex tracking systems, but does have the limitation that the user must be facing the camera and not blocking the camera's view of the ball. In addition, accuracy in the depth dimension is worse than in the horizontal and vertical dimensions.

Probably the best candidate for self-contained 6-DOF tracking is inside-out vision-based tracking, in which the tracked object uses a camera to view the world, and analyzes the changes in this view over time to understand its own motion (translations and rotations). Although this approach is inherently relative, such systems can keep track of "feature points" in the scene to give a sort of absolute tracking in a fixed coordinate system connected with the scene. Algorithms such as parallel tracking and mapping (PTAM) (Klein & Murray, 2007) are getting closer to making this a reality.

Three recent tracking developments deserve special mention, as they are bringing many new designers and researchers into the realm of 3D UIs. The first is the Nintendo Wii Remote. This gaming peripheral does not offer 6-DOF tracking, but does include several inertial sensors in addition to a simple optical tracker that can be used to move a cursor on the screen. Wingrave and colleagues (Wingrave et al, 2010) presented a nice discussion of how the Wii Remote differs from traditional trackers, and how it can be used in 3D UIs.

Second, the Microsoft Kinect (Figure 32.1) delivers tracking in a very different way. Rather than tracking a handheld device or a single point on the user's head, it uses a depth camera to track the user's entire body (a skeleton of about 20 points). The 3-DOF position of each point is measured, but orientation is not detected. And since it tracks the body directly, no "controller" is needed. Researchers have designed some interesting 3D interactions with Kinect (e.g., Wilson & Benko, 2010), but they are necessarily quite different than those based on single-point 6-DOF tracking.

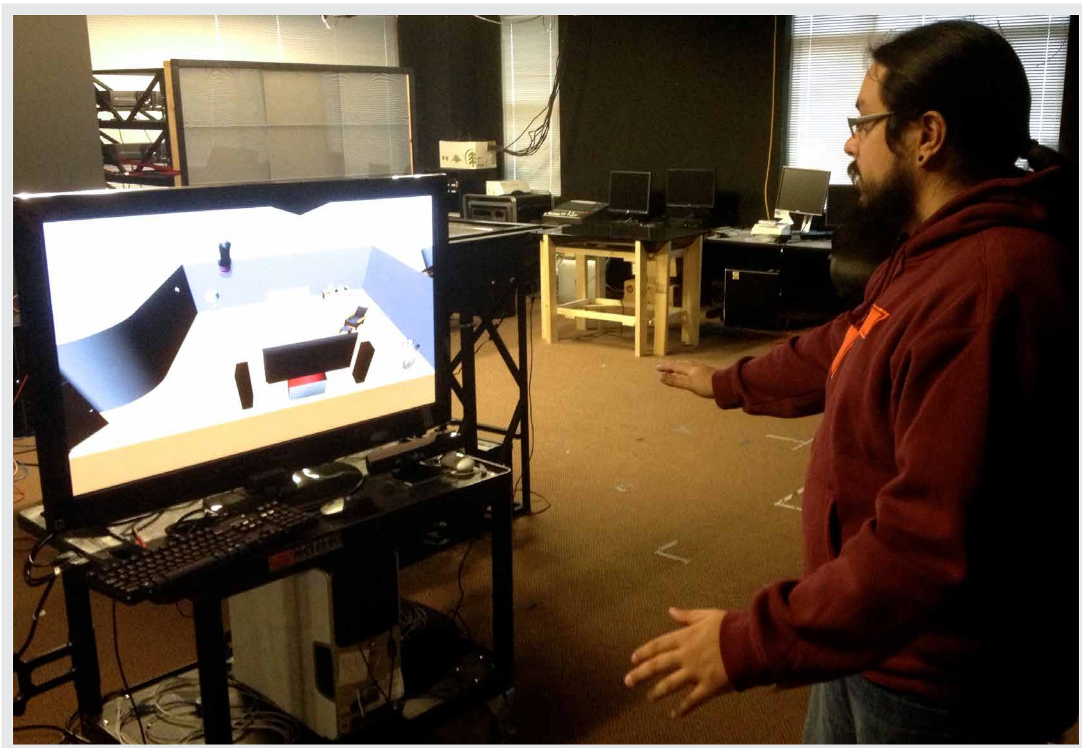


Figure 32.1: 3D interaction with Microsoft Kinect.



Third, the Leap Motion device, which has been announced but is not available at the time of this writing, promises to deliver very precise 3D tracking of hands, fingers, and tools in a small workspace. It has the potential to make 3D interaction a standard part of the desktop computing experience, but we will have to wait and see how best to design interaction techniques for this device. It will share many of the benefits and drawbacks of the Kinect, and although it is designed to support "natural" interaction, naturalism is not always possible, and not always the best solution (as we will discuss below).

For 3D interaction, spatial trackers are most often used inside handheld devices. These devices typically include other inputs such as buttons, joysticks, or trackballs, making them something like a "3D mouse." Like desktop mice, these can then be used for pointing, manipulating objects, selecting menu items, and the like. Trackers are also used to measure the user's head position and orientation. Head tracking is useful for modifying the view of a 3D environment in a natural way.

The type of spatial tracker used in a 3D UI can have a major impact on its usability, and different trackers may require different UI designs. For example, a tracker with higher latency might not be appropriate for precise object manipulation tasks, and an interface using a 3-DOF orientation tracker requires additional methods for translating the viewpoint in the 3D environment, since it does not track the user's position.

This short section can't do justice to the complex topic of spatial tracking. An older, but very good, overview of tracking technologies and issues can be found in Welch's paper (Welch & Foxlin, 2002).

### 32.3.2 Other Input Devices

While spatial tracking is the fundamental input device for 3D UIs, it is usually not sufficient on its own. As noted above, most handheld trackers include other sorts of input, because it's difficult to map all interface actions to position, orientation, or motion of the tracker. For example, to confirm a selection action, a discrete event or command is needed, and a button is much more appropriate for this than a hand motion. The Intersense IS900 wand is typical of such handheld trackers; it includes four standard buttons, a "trigger" button, and a 2-DOF analog joystick (which is also a button) in a handheld form factor. The Kinect, because of its "controller-less" design, suffers from the lack of discrete inputs such as buttons.

Generalizing this idea, we can see that almost any sort of input device can be made into a spatial input device by tracking it. Usually this requires adding some hardware to the device, such as optical tracking markers. This extends the capability and expressiveness of the tracker, and allows the input from the device to be interpreted differently depending on its position and orientation. For example, in my lab we have experimented with tracking multi-touch smartphones and combining the multi-touch input with the spatial input for complex object manipulation interfaces (Wilkes et al., 2012). Other interesting devices, such as bend-sensitive tape, can be tracked to provide additional degrees of freedom (Balakrishnan et al., 1999).

Gloves (or finger trackers) are another type of input device that is frequently combined with spatial trackers. Pinch gloves detect contacts between the fingers, while data gloves and finger trackers measure joint angles of the fingers. Combining these with trackers allow for interesting, natural, and expressive use of hand gestures, such as in-air typing (Bowman et al., 2002), writing (Ni et al., 2011), or sign language input (Fels & Hinton, 1997).

### 32.3.3 Displays

Much of the early work on 3D UIs was done in the context of interaction with VR systems, which use some form of "immersive" display, such as head-mounted displays (HMDs), surround-screen displays (e.g., CAVEs), or wall-sized stereoscopic displays. Increasingly, however, 3D interaction is taking place with TVs or even desktop monitors, due to the use of consumer-level tracking devices meant for gaming. Differences in display configuration and characteristics can have a major impact on the design and usability of 3D UIs.

HMDs (Figure 32.2) provide a full 360-degree surround (when combined with head tracking) and can block out the user's view of the real world, or enhance the view of the real world when used in AR systems. When used for VR, HMDs keep users from seeing their own hands or other parts of their bodies, meaning that devices must be usable eyes-free, and that users may be hesitant to move around in the physical environment. HMDs also vary widely in field of view (FOV). When a low FOV is present, 3D UI designers must use the limited screen real estate sparingly.

FIGURE 32.2: Using a 3D UI while wearing a head-mounted display. The TV in the background shows the image displayed in the HMD.



CAVE-like displays (Cruz-Neira et al., 1993) may provide a full surround, but more often use two to four screens to partially surround the user. Among other considerations, for 3D UIs this means that the designer must provide a way for the user to rotate the world. The mixture of physical and virtual viewpoint rotation can be confusing and can reduce performance on tasks like visual search (McMahan, 2011).

3D UIs on smaller displays like TVs also pose some interesting challenges. With HMDs and CAVEs, the software field of view (the FOV of the virtual camera) is usually matched to the physical FOV of the display so that the view is realistic, as if looking through a window to the virtual world. With desktop monitors and TVs, however, we may not know the size of the display or the user's position relative to it, so determining the appropriate software FOV is difficult. This in turn may influence the user's ability to understand the scale of objects being displayed.

Finally, we know that display characteristics can affect 3D interaction performance. Prior research in my lab has shown, for example, that stereoscopic display can improve performance on difficult manipulation tasks (Narayan et al., 2005) but not on simpler manipulation tasks (McMahan et al., 2006).

## 32.4 DESIGNING USABLE 3D UIS

As a serious topic in HCI, 3D interaction has not been around very long. The seminal papers in the field were only written in the mid- to late-1990s, the most-cited book in the field was published in 2005, and the IEEE Symposium on 3D User Interfaces didn't begin until 2006.

Because of this, the level of maturity of 3D UI design principles lags behind those for standard GUIs. There is no standard 3D UI (and it's not clear that there could be, given the diversity of input devices, displays, and interaction techniques), and few well-established guidelines for 3D UI design. While general HCI principles such as Nielsen's heuristics (Nielsen & Molich, 1990) still apply, they are not sufficient for understanding how to design a usable 3D UI.

Thus, it's important to have specific design principles for 3D interaction. While the 3D UI book (Bowman et al., 2005) and several other works (Kulik, 2009; Gabbard, 1997; Kaur, 1999) have extensive lists of guidelines, here I've tried to distill what I feel are the most important lessons about good 3D UI design.

### 32.4.1 Understand the design space

Despite the youth of the field, there is a very large number of existing 3D interaction techniques for the so-called "universal tasks" of travel, selection, manipulation, and system control. In many cases, these techniques can be reused directly or with slight modifications in new applications. The lists of techniques in the 3D

UI book (Bowman et al., 2005) are a good place to start; more recent techniques can be found in the proceedings of IEEE 3DUI and VR, ACM CHI and UIST, and other major conferences.

When existing techniques are not sufficient, new techniques can sometimes be generated by combining existing technique components. Taxonomies of technique components (Bowman et al., 2001) can be used as design spaces for this purpose.

### 32.4.2 There is still room to innovate

a wide variety of techniques already existsit is impossible to innovate in 3D UI design. On one hand, most of the primary metaphors for the universal tasks have probably been invented already. On the other hand, there are several reasons to believe that there are new, radically different metaphors than what we currently have.

First, we know the design space of 3D interaction is very large due to the number of devices and mappings available. Second, 3D interaction design can be magical—limited only by the designer's imagination. Third, new technologies (such as the Leap Motion device) with the potential for new forms of interaction are constantly appearing. For example, in a recent project in our lab, students used a combination of recent technologies (multi-touch tablet, 3D reconstruction, marker-based AR tracking, and stretch sensors) to enable "AR Angry Birds"—a novel form of physical interaction with both real and virtual objects in AR (Figure 32.3). Finally, techniques can be designed specifically for specialized tasks in various application domains. For example, we designed domain-specific interaction techniques for object cloning in the architecture and construction domain (Chen and Bowman, 2009).

Figure 32.3: AR Angry Birds prototype with a physical slingshot and "destruction" of real-world objects—an example of innovation in 3D UI design.



### 32.4.3 Be careful with mappings and DOFs

One of the most common problems in 3D UI design is the use of inappropriate mappings between input devices and actions in the interface. Zhai & Milgram (1993) showed, for instance, that elastic sensors (e.g., a joystick) and isometric sensors (e.g., a SpaceBall) map well to rate-controlled movements, where the displacement or force measured by the sensor is mapped to velocity of an object (including the viewpoint) in the virtual world, while isotonic sensors (e.g., a position tracker) map well to position-controlled movements, where the position measured by the sensor is mapped to the position of an object. When this principle is violated, performance suffers.

Similarly, there are often problems with the mappings of input DOFs to actions. When a high-DOF input is used for a task that requires a lower number of DOFs, task performance can be unnecessarily difficult. For example, selecting a menu item is inherently a one-dimensional task. If users need to position their virtual hands within a menu item to select it (a 3-DOF input), the interface requires too much effort.

Another DOF problem is the misuse of integral and separable DOFs. Jacob & Sibert (1992) showed that input devices with integral DOFs (those that are controlled all together, as in a 6-DOF tracker) should be mapped to tasks that users perceive as integral (such as 6-DOF object manipulation), while input devices with separable DOFs (those that can be controlled independently, such as a set of sliders) should be mapped to tasks that have sub-tasks users perceive as separable (such as setting the hue, saturation, and value of a color). A violation of this concept, for example, would be to use a six-DOF tracker to simultaneously control the 3D position of an object and the volume of an audio clip, since those tasks cannot be integrated by the user.

In general, 3D UI designers should seek to reduce the number of DOFs the user is required to control. This can be done by using lower-DOF input devices, by ignoring some of the input DOFs, or by using physical or virtual constraints. For example, placing a virtual 2D interface on a physical tablet prop (Schmalstieg et al., 1999) provides a constraint allowing users to easily use 6-DOF tracker input for 2D interaction.

### 32.4.4 Keep it simple

Although 3D UIs can be very expressive and can support complex tasks, not all tasks in a 3D UI need to use fully general interaction techniques. When the user's goal is simple, designers should provide simple and effortless techniques. For example, there are many general-purpose travel techniques that allow users to control the position and orientation of the viewpoint continuously, but if the user simply wants to move to a known landmark, a simple target-based technique (e.g., point at the landmark object) will be much more usable.

Reducing the number of DOFs, as described above, is another way to simplify 3D UIs. For instance, travel techniques can require only two DOFs if terrain following is enabled.

Finally, when using physical buttons or gestures to map to commands/functions, avoid the tendency to add another button or gesture for each new command. Users typically can't remember a large number of gestures, and remembering the mapping between buttons and functions becomes difficult after only 2-3 buttons are used.

### 32.4.5 Design for the hardware

In traditional UIs, we usually try to design without regard for the display or the input device (i.e., display- and device-independence). UIs should be just as usable no matter whether you are using a large monitor or a small laptop, with a mouse or a trackpad. This is not always strictly true—when you have a very large multi-monitor setup, for example. But in 3D UIs, what works on one display or with one device very rarely works exactly the same way on different systems.

We call this the *migration* issue. When migrating to a different display or device, the UI and interaction techniques often need to be modified. In other words, we need display- and device-specific 3D UIs.

For example, the World-in-Miniature (WIM) technique (Stoakley et al., 1995), which allows users to move virtual objects in a full-scale virtual environment by manipulating small "dollhouse" representations of those objects, was originally designed for an HMD with two handheld trackers for input. When we tried to migrate WIM to a CAVE (Bowman et al., 2007), we found performance to be significantly worse, probably because users found it difficult to fuse the stereo imagery when the virtual WIM was held close to their eyes. In addition, we had to add controls for rotating the world due to the missing back wall of the CAVE. More recently, we tried to migrate WIM to use the Kinect, and were not able to find *any* reasonable mapping that allowed users to easily manipulate both the WIM and the virtual hand with six DOFs.

### 32.4.6 You may still have to train users, but a little training can go a long way

3D interaction is often thought of as "natural," but for many novice users, effective operation of 3D UIs is anything but natural. Users in HMDs don't want to turn their heads, much less move their bodies. Moving a hand in two dimensions (parallel to a screen) is fine, but moving a hand towards or away from the screen doesn't come naturally. When using 3D travel techniques, users don't take advantage of the ability to fly, or to move sideways, or to walk through virtual walls (Bowman et al., 1999).

Because of this, we find that we often have to train our users before they become proficient at using even well designed 3D UIs. In most of the HCI community, the need for training or instruction is seen as a sign of bad design, but in the examples mentioned above, effective use requires users to go against their instincts and intuitions. If a minimal (one-minute) training session allows users to improve their performance significantly, we see that as both practical and positive.

### 32.4.7 Always evaluate

Finally, we suggest that all 3D UI designs should undergo formative, empirical usability evaluation with members of the target user population. While this guideline probably applies to all UIs, 3D UIs in particular are difficult to design well based on theory, principles, and intuition alone. Many usability problems don't become clear until users try the 3D UI. evaluate early and often.

## 32.5 CURRENT 3D UI RESEARCH

In this final section, I want to highlight two of the interesting problems 3D UI researchers are addressing today.

### 32.5.1 Realism vs. Magic - The Question of Interaction Fidelity

One of the fundamental issues in 3D UI design is the tension between realistic and magical interaction. Many feel that 3D interaction should be as "natural" as possible, reusing and reproducing interactions from the real world so that users can take advantage of their existing skills, knowing what to do and how to do it. On the other hand, 3D UIs primarily allow users to interact with virtual objects and environments, whose only constraints are due to the skill of the programmer and the limits of the technology. Thus, "magic" interaction is possible, enabling the user to transcend the limitations of human perception and action, to reduce or eliminate the need for physical effort and lengthy operations, and even to perform tasks that are impossible in the real world.

This question is related to the concept of *interaction fidelity*, which we define as the objective degree with which the actions (characterized by movements, forces, body parts in use, etc.) used for a task in the UI correspond to the actions used for that task in the real world (Bowman et al., 2012). By talking about the *degree* of fidelity, we emphasize that we are not just talking about "realistic" and "non-realistic" interactions, but a continuum of realism, which itself has several different dimensions.

Consider an example. For the task of moving a virtual book from one location on a desk to another, we could, among many other options: a) map the movements of the user's real hand and fingers exactly, requiring exact placement, grasping, and releasing, b) position a 3D cursor over the book, press a button, move the cursor to the target position, and release the button, or c) choose "move" from a menu, and then use a laser pointer to indicate the book and the target location. Clearly, option a) is the most natural, option b) uses a natural metaphor but leaves out some of the less necessary details of the real-world interaction, and option c) has very low interaction fidelity. Option a) is probably the easiest for a novice user to learn and use, providing that the designer can replicate the actions and perceptual cues from the real world well enough, although option b) is the simplest and may be just as effective.

Some tasks are very difficult (or impossible) to do in the real world. What if I want to remove a building from a city? A highly natural 3D UI would require the user to obtain some virtual explosives or a virtual crane with a wrecking ball, and operate these over a long period of time. Here a "magic" technique, such as allowing the user to "erase" the building, or selecting the building and invoking a "delete" command by voice, is clearly more practical and effective.

In many cases of difficult tasks, the question is not whether we should use a natural or magical 3D UI, because the purely natural technique wouldn't be practical. Instead, the question is whether to use a natural *metaphor*. For example, in the real world I cannot pick up objects that are beyond arm's reach, but in the virtual world I can. Should I do this with a reaching and grasping metaphor, as in the Go-Go technique (Poupyrev et al., 1996), which extends the user's virtual hand far into the environment based on natural movements? Or should I pick up the object by pointing to it using a laser pointer metaphor, as in the HOMER technique

(Bowman & Hodges, 1997)? In this case, the less natural laser pointer metaphor is more effective in terms of user performance, but enhanced natural metaphors are easy to learn and highly usable in many situations.

Because techniques like Go-Go use natural metaphors to extend users' abilities beyond what's possible in the real world, we refer to them as *hyper-natural*. There is not a single answer to the question of whether to choose natural, hyper-natural, or non-natural magic techniques, but overall, research has shown significant benefits for the natural and hyper-natural design approaches (Bowman et al., 2012).

### 32.5.2 Increasing Precision

A major disadvantage of 3D UIs based on spatial tracking systems is the difficulty of providing precise 3D spatial input. The modern mouse is a highly precise, accurate, and responsive 2D spatial input device—users can point at on-screen elements, even individual pixels, quickly and accurately. 3D spatial tracking systems are far behind the mouse in terms of precision (jitter), accuracy of reported values, and responsiveness (latency), making it problematic to use them for tasks requiring precision (Teather et al., 2009).

But even if 3D spatial tracking systems improve their specifications to be comparable with today's mouse, 3D UIs will still have a precision problem, for the following reasons:

3D interaction is performed in the air, not on a surface. There is no friction or physical support to make movements more controlled and precise.

Humans have a natural hand tremor that causes in-air movements to be jittery.

Interfaces based on 3D pointing using ray-casting (i.e., laser pointer metaphor) amplify this hand tremor so that it becomes worse the farther out along the ray you go.

3D spatial trackers are not "parkable" like the mouse—the user cannot let go of them and be assured that they will stay in the same position.

So is there any hope of 3D UIs that can be used for precise work? A partial solution is to filter the output of 3D spatial trackers to reduce noise, but filtering can cause other problems, such as increased latency. Current research is addressing the precision problem using several different strategies.

One approach is to modify the control/display (C/D) ratio. The simple idea here is to use an N:1 mapping between movements of the input device (control) and movements in the system (display), where N is greater than one. In other words, if the C/D ratio is five, then a five-centimeter movement (or five-degree rotation) of the tracker would result in a one-centimeter movement (or one-degree rotation) in the virtual world. This gives users greater levels of control and the ability to achieve more precision, but at the cost of increased physical effort and time. Some techniques (e.g., Frees et al., 2007) dynamically modify the C/D ratio so that precision is only added when necessary (e.g., when the user is moving slowly).

A second strategy is to ensure that the user is not required to be more precise than absolutely necessary. For example, if the user is selecting a very small object in a sparse environment, there is no need to make the user touch or point to the object precisely. Rather, the cursor can have area or volume (e.g., a circle or sphere) instead of being a point (e.g., Liang & Green, 1994), or the cursor can snap to the nearest object (e.g., de Haan et al., 2005).

Finally, a promising approach called *progressive refinement* spreads out the interaction over time rather than requiring a single precise action. A series of rough, imprecise actions can be used to achieve a precise result, without a great deal of effort on the part of the user. For instance, the SQUAD technique (Kopper et al., 2011) allows users to select small objects in cluttered environments by first doing a volume selection, then refining the set of selected objects with a series of rapid menu selections. In very difficult cases, this technique was even faster than ray-casting, which uses a single precise selection, and in all cases, SQUAD resulted in fewer selection errors. This progressive refinement approach should be broadly applicable to many sorts of difficult 3D interaction tasks.

## 32.6 FOR FURTHER READING

For an overview of the field of 3D UIs, and a comprehensive survey of devices and interaction techniques, see *3D User Interfaces: Theory and Practice* (Bowman et al., 2005).

The best current research in the field can be found in the proceedings of the *IEEE Symposium on 3D User Interfaces*.

For more on how to use realism and magic in 3D UI design, see a recent tutorial in *IEEE Computer Graphics & Applications* (Kulik, 2009).

Wolfgang Stuerzlinger provides a set of practical guidelines from his years of experience in 3D UI design in a recent survey paper (Bowman et al., 2008).

To learn more about experimental results on the effects of interaction fidelity in 3D UIs, see my recent *Communications of the ACM* paper (Bowman et al., 2012).

## 32.7 REFERENCES

*Ashley, Steven (2008): Annotating the Real World: Augmented Reality Makes Commercial Headway. In Scientific American, 299 (4) pp. 27-28*

*Balakrishnan, Ravin, Fitzmaurice, George W., Kurtenbach, Gordon and Singh, Karan (1999): Exploring interactive curve and surface manipulation using a bend and twist sensitive input strip. In: SI3D 1999 1999. pp. 111-118*

*Bowman, Doug A. and Hodges, Larry F. (1997): An Evaluation of Techniques for Grabbing and Manipulating Remote Objects in Immersive Virtual Environments. In: SI3D 1997 1997. pp. 35-38,182*

*Bowman, Doug A., McMahan, Ryan P. and Ragan, Eric D. (2012): Questioning naturalism in 3D user interfaces. In Communications of the ACM, 55 (9) pp. 78-88*

*Bowman, Doug A., Johnson, Donald B. and Hodges, Larry F. (2001): Testbed Evaluation of Virtual Environment Interaction Techniques. In Presence: Teleoperators and Virtual Environments, 10 (1) pp. 75-95*

*Bowman, Doug A., Kruijff, Ernst, LaViola, Joseph J. and Poupyrev, Ivan (2005): 3D User Interfaces: Theory and Practice. Addison-Wesley Professional*

*Bowman, Doug A., Wingrave, Chadwick A., Campbell, J. M., Ly, V. Q. and Rhoton, C. J. (2002): Novel Uses of Pinch GlovesTM for Virtual Environment Interaction Techniques. In Virtual Reality, 6 (3) pp. 122-129*

*Bowman, Doug A., Davis, Elizabeth Thorpe, Hodges, Larry F. and Badre, Albert N. (1999): Maintaining Spatial Orientation during Travel in an Immersive Virtual Environment. In Presence: Teleoperators and Virtual Environments, 8 (6) pp. 618-631*

*Bowman, Doug A., Coquillart, Sabine, Froehlich, Bernd, Hirose, Michitaka, Kitamura, Yoshifumi, Kiyokawa, Kiyoshi and Stürzlinger, Wolfgang (2008): 3D User Interfaces: New Directions and Perspectives. In IEEE Computer Graphics and Applications, 28 (6) pp. 20-36*

*Bowman, Doug A., Badillo, Brian and Manek, Dhruv (2007): Evaluating the Need for Display-Specific and Device-Specific 3D Interaction Techniques. In: Shumaker, Randall (ed.) ICVR 2007 - Virtual Reality - Second International Conference - Part 1 July 22-27, 2007, Beijing, China. pp. 195-204*

*Chen, Jian and Bowman, Doug A. (2009): Domain-Specific Design of 3D Interaction Techniques: An Approach for Designing Useful Virtual Environment Applications. In Presence: Teleoperators and Virtual Environments, 18 (5) pp. 370-386*

*Cruz-Neira, Carolina, Sandin, Daniel J. and DeFanti, Thomas A. (1993): Surround-screen projection-based virtual reality: the design and implementation of the CAVE. In: Proceedings of the 20st Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH 1993 1993. pp. 135-142*

*Feiner, Steven K., MacIntyre, Blair and Seligmann, Doree Duncan (1992): Annotating the real world with knowledge--based graphics on a see--through head--mounted display. In: Graphics Interface 92 May 11-15, 1992, Vancouver, British Columbia, Canada. pp. 78-85*

*Fels, S. S. and Hinton, G. E. (1997): Glove-talk II - a neural-network interface which maps gestures to parallel formant speech synthesizer controls. In IEEE Transactions on Neural Networks, 8 (5) pp. 977-984*

*Frees, Scott, Kessler, G. Drew and Kay, Edwin (2007): PRISM interaction for enhancing control in immersive virtual environments. In ACM Transactions on Computer-Human Interaction, 14 (1) p. 2*

*Gabbard, Joseph L. (1997). Taxonomy of Usability Characteristics in Virtual Environments. (M.S. Thesis). Virginia Tech.*

*Google (2012). Project Glass. Retrieved 7 November 2012 from Google: https://plus.google.com/+projectglass/posts.*

*Haan, Gerwin de, Koutek, Michal and Post, Frits H. (2005): IntenSelect: Using Dynamic Object Rating for Assisting 3D Object Selection. In: Kjems, Eric and Blach, Roland (eds.) Proceedings of the 9th Int. Workshop on Immersive Projection Technology - 11th Eurographics Workshop on Virtual Environments - IPT-EGVE 2005 2005, Aalborg, Denmark. pp. 201-209*

*Höllerer, Tobias, Feiner, Steven, Terauchi, Tachio, Rashid, Gus and Hallaway, Drexel (1999): Exploring MARS: developing indoor and outdoor user interfaces to a mobile augmented reality system. In Computers & Graphics, 23 (6) pp. 779-785*

*Jacob, Robert J. K. and Sibert, Linda E. (1992): The Perceptual Structure of Multidimensional Input Device Selection. In: Bauersfeld, Penny, Bennett, John and Lynch, Gene (eds.) Proceedings of the ACM CHI 92 Human Factors in Computing Systems Conference June 3-7, 1992, Monterey, California. pp. 211-218*

*Kaur, K. (1999). Designing Virtual Environments for Usability. Doctoral Dissertation. University College, London*

*Klein, Georg and Murray, David W. (2007): Parallel Tracking and Mapping for Small AR Workspaces. In: Sixth IEEE/ACM International Symposium on Mixed and Augmented Reality, ISMAR 2007, 13-16 November 2007, Nara, Japan 2007. pp. 225-234*

*Kopper, Regis, Bowman, Doug A., Silva, Mara G. and McMahan, Ryan P. (2010): A human motor behavior model for distal pointing tasks. In International Journal of Human-Computer Studies, 68 (10) pp. 603-615*

*Kopper, Regis, Bacim, Felipe and Bowman, Doug A. (2011): Rapid and accurate 3D selection by progressive refinement. In: Proceedings of the 2011 IEEE Symposium on 3D User Interfaces 2011. pp. 67-74*

*Kulik, Alexander (2009): Building on Realism and Magic for Designing 3D Interaction Techniques. In IEEE Computer Graphics and Applications, 29 (6) pp. 22-33*

*LaViola, Joseph J. (2008): Bringing VR and Spatial 3D Interaction to the Masses through Video Games. In IEEE Computer Graphics and Applications, 28 (5) pp. 10-15*

*Liang, Jiandong and Green, Mark (1994): JDCAD: A highly interactive 3D modeling system. In Computers & Graphics, 18 (4) pp. 499-506*

*McMahan, Ryan Patrick (2011). Exploring the Effects of Higher-Fidelity Display and Interaction for Virtual Reality Games. (Ph.D. dissertation). Virginia Tech.*

*McMahan, Ryan P., Gorton, Doug, Gresock, Joe, McConnell, Will and Bowman, Doug A. (2006): Separating the effects of level of immersion and 3D interaction techniques. In: Slater, Mel, Kitamura, Yoshifumi, Tal, Ayellet, Amditis, Angelos and Chrysanthou, Yiorgos (eds.) VRST 2006 - Proceedings of the ACM Symposium on Virtual Reality Software and Technology November 1-3, 2006, Limassol, Cyprus. pp. 108-111*

*Meyer, Kenneth, Applewhite, Hugh L. and Biocca, Frank A. (1992): A Survey of Position Trackers. In Presence: Teleoperators and Virtual Environments, 1 (2) pp. 173-200*

*Mistry, Pranav and Maes, Pattie (2009): SixthSense: a wearable gestural interface. In: Oda, Yuko and Tanaka, Mariko (eds.) International Conference on Computer Graphics and Interactive Techniques, SIGGRAPH ASIA 2009, Yokohama, Japan, December 16-19, 2009, Art Gallery and Emerging Technologies Adaptation 2009. p. 85*

*Narayan, Michael, Waugh, Leo, Zhang, Xiaoyu, Bafna, Pradyut and Bowman, Doug A. (2005): Quantifying the benefits of immersion for collaboration in virtual environments. In: Singh, Gurminder, Lau, Rynson W. H., Chrysanthou, Yiorgos and Darken, Rudolph P. (eds.) VRST 2005 - Proceedings of the ACM Symposium on Virtual Reality Software and Technology November 7-9, 2005, Monterey, CA, USA. pp. 78-81*

*Ni, Tao, Bowman, Doug A. and North, Chris (2011): AirStroke: bringing unistroke text entry to freehand gesture interfaces. In: Proceedings of ACM CHI 2011 Conference on Human Factors in Computing Systems 2011. pp. 2473-2476*

*Nielsen, Jakob and Molich, Rolf (1990): Heuristic evaluation of user interfaces. In: Carrasco, Jane and Whiteside, John (eds.) Proceedings of the ACM CHI 90 Human Factors in Computing Systems Conference 1990, Seattle, Washington,USA. pp. 249-256*

*Poupyrev, Ivan, Billinghurst, Mark, Weghorst, Suzanne and Ichikawa, Tadao (1996): The Go-Go Interaction Technique: Non-Linear Mapping for Direct Manipulation in VR. In: Kurlander, David, Brown, Marc and Rao, Ramana (eds.) Proceedings of the 9th annual ACM symposium on User interface software and technology November 06 - 08, 1996, Seattle, Washington, United States. pp. 79-80*

*Rekimoto, Jun (1998): Multiple-Computer User Interfaces: A Cooperative Environment Consisting of Multiple Digital Devices. In: Streitz, Norbert A., Konomi, Shin'ichi and Bur-*

*khardt, Heinz Jürgen (eds.) Cooperative Buildings, Integrating Information, Organization, and Architecture, First International Workshop, CoBuild98, Darmstadt, Germany, February 1998, Proceedings 1998. pp. 33-40*

*Schmalstieg, Dieter, Encarnacao, L. Miguel and Szalavári, Zsolt (1999): Using transparent props for interaction with the virtual table. In: SI3D 1999 1999. pp. 147-153*

*Stoakley, Richard, Conway, Matthew and Pausch, Randy (1995): Virtual Reality on a WIM: Interactive Worlds in Miniature. In: Katz, Irvin R., Mack, Robert L., Marks, Linn, Rosson, Mary Beth and Nielsen, Jakob (eds.) Proceedings of the ACM CHI 95 Human Factors in Computing Systems Conference May 7-11, 1995, Denver, Colorado. pp. 265-272*

*Teather, Robert J., Pavlovych, Andriy, Stuerzlinger, Wolfgang and MacKenzie, I. Scott (2009): Effects of tracking technology, latency, and spatial jitter on object movement. In: Proceedings of the 2009 IEEE Symposium on 3D User Interfaces 2009. pp. 43-50*

*van Dam, Andries (1997): Post-WIMP User Interfaces. In Communications of the ACM, 40 (2) pp. 63-67*

*Vogel, Daniel and Balakrishnan, Ravin (2005): Distant freehand pointing and clicking on very large, high resolution displays. In: Proceedings of the 2005 ACM Symposium on User Interface Software and Technology 2005. pp. 33-42*

*Welch, Greg and Foxlin, Eric (2002): Motion Tracking: No Silver Bullet, but a Respectable Arsenal. In IEEE Computer Graphics and Applications, 22 (6) pp. 24-38*

*Wigdor, Daniel and Wixon, Dennis (2011): Brave NUI World: Designing Natural User Interfaces for Touch and Gesture. Morgan Kaufmann*

*Wilkes, Curtis B., Tilden, Dan and Bowman, Doug A. (2012): 3D User Interfaces Using Tracked Multi-touch Mobile Devices. In: Joint Virtual Reality Conference of ICAT - EGVE - EuroVR 2012 2012.*

*Wilson, Andrew D. and Benko, Hrvoje (2010): Combining multiple depth cameras and projectors for interactions on, above and between surfaces. In: Proceedings of the 2010 ACM Symposium on User Interface Software and Technology 2010. pp. 273-282*

*Wingrave, Chadwick A., Williamson, Brian, Varcholik, Paul, Rose, Jeremy, Miller, Andrew, Charbonneau, Emiko, Bott, Jared N. and LaViola, Joseph J. (2010): The Wiimote and Beyond: Spatially Convenient Devices for 3D User Interfaces. In IEEE Computer Graphics and Applications, 30 (2) pp. 71-85*

*Zhai, Shumin and Milgram, Paul (1993): Human Performance Evaluation of Manipulation Schemes in Virtual Environments. In: VR 1993 1993. pp. 155-161*

## About the author

### Doug A. Bowman

© *Doug A. Bowman*

Has also published under the name of:
"Doug Bowman"

Personal Homepage:
https://research.cs.vt.edu/3di/user/123

Doug A. Bowman is a Professor of Computer Science and Director of the Center for Human-Computer Interaction at Virginia Tech. He is the principal investigator of the 3D Interaction Group, focusing on the topics of three-dimensional user interface design and the benefits of immersion in virtual environments. Dr. Bowman was the lead author of 3D User Interfaces: Theory and Practice, and served as the general chair of the IEEE Virtual Reality Conference in 2007-2008. He received a CAREER award from the National Science Foundation for his work on 3D Interaction, and has been named an ACM Distinguished Scientist. His undergraduate degree in mathematics and computer science is from Emory University, and he received his M.S. and Ph.D. in computer science from the Georgia Institute of Technology.

## Your notes and thoughts on chapter 32

Record your notes and thoughts on this chapter. If you want to share these thoughts with others online, go to the bottom of the page at:
http://www.interaction-design.org/encyclopedia/3d_user_interfaces.html

**NOTES:**

# CHAPTER 33

# Action Research

## Its Nature and Relationship to Human-Computer Interaction

*by Ned Kock.*

Technology and action are two elements that define what it is to be human. It is technology that has made *Homo sapiens* such a successful species, and it is the actions enabled by technology that will ensure that we continue to be successful à or meet our doom. No technological development has any value without action. Action Research (AR) is all about action and, at the same time, it is also rigorous research. When applied in the context of technology, Action Research is the study of how technology is applied in the real world and the practical consequences of technology-enabled action.

More broadly, Action Research is a generic name used to refer to a set of research approaches that share a few common characteristics. In Action Research the researcher typically tries to provide a service to a research "client", often an organization, and at the same time add to the body of knowledge in a particular domain.

In technology-related inquiry, an Action Research study could entail the researcher introducing a new technology in an organization, and at the same time studying the effects of the technology in that organization. The organization can be an insurance company, a church, a distributed film-editing organization, or an online mutual support community of diabetics. The emphasis may be on technology design, empirical evaluation of the effects of one ore more technologies, or both. The emphasis of my discussion in this encyclopedia entry is on the empirical evaluation modality, because that is the modality that is the most closely related to Action Research. One may design a unique technology, but the technology will only become useful when it is employed in practice; Action Research is all about practice.

## 33.1 ACTION RESEARCH

According to most accounts, Action Research originated independently in the U.S.A. and England in the 1940s. In the U.S.A., Action Research emerged from the work of Kurt Lewin on a variety of topics, ranging from child welfare to group dynamics. Lewin was a German-born social psychologist whom many see as the "father" of Action Research. In England, Action Research's origins are not tied to a particular individual, but to an institution û the Tavistock Institute of Human Relations in London. There Action Research was used as a research method to both understand and treat socio-psychological disorders associated with war-related experiences.

To say that the range of areas and ways in which Action Research can be conducted is vast is an understatement. Action Research can be used in many general fields of inquiry such as bilingual education, clinical psychology, sociology, and information systems. It can be conducted in ways that are aligned with most epistemologies, including the positivist, interpretivist, and critical epistemologies. Action Research can have as its unit of analysis the individual, the small group, and even the entire organization. It can be used to address issues as varied as health concerns and environmental problems, and evaluate the impact of things like engineering techniques and business methods.

One of the key characteristics that distinguishes Action Research from most other research approaches, and also constitutes one of its main appeals, is that Action Research aims at both improving the subject of the study (often called the research "client"), and generating knowledge, achieving both *at the same time*. While this characteristic may seem straightforward enough to easily differentiate Action Research from most other research approaches û such as experimental, survey, and case research û it is not. Action Research will be contrasted with those other research approaches in the next section.

Let us assume, for the sake of illustration, that a survey-based research project was conducted addressing the differential access to the Internet between two main income groups, one high (wealthy) and the other low (poor), in a particular city, where the reasons for the digital divide are unclear. Can that research be considered Action Research if a report based on it is used by the city's government to bridge the gap that characterizes the divide? The answer is "yes", if the research encompasses the city's actions, and possibly a follow-up survey assessment of the impact of those actions. The answer is "no", if the research ended with the analysis of the survey and the publication of the summary report.

Because of Action Research's dual goal, researchers employing it are said to have to satisfy two "masters" û the subject (or subjects) of the research, and the research community. Historically, one could argue that it has been harder to satisfy the latter, especially in fields of inquiry where Action Research has not traditionally been used very often, such as in technology-related research.

## 33.2 ACTION RESEARCH VIS-A-VIS OTHER RESEARCH APPROACHES

The literature on empirical research methods suggests that three general approaches have accounted for most of the published empirical investigations on the effects of technologies on individuals in organizations. These three general research approaches are: experimental, survey, and case research. The paragraphs below provide a brief description of these three research approaches, and contrast them with Action Research.

### 33.2.1 Experimental research

This approach has its roots in the scientific practice of biologists, physicists, and medical professionals. Variables are manipulated over time, associated numeric data is collected, and causal or correlation models are tested through inferential statistical analysis procedures. The researcher has a strong control over the environment being observed. Experimental research is typically applied to test models or hypotheses. An example would be a research study in which a group of people used a technology (e.g., an online collaborative writing tool) to perform a group-based task (e.g., collaborative writing of a sales contract), and another group performed the task by interacting face-to-face. The technology group, in this case, would be called the "treatment" group, while the other would be the "control" group. Both groups would be taken randomly from a larger group of people. An example of hypothesis to be tested in this scenario could be that the technology group would write a "better" contract than the no-technology group.

### 33.2.2 Survey research

This approach originates from the work of economists and sociologists. The researcher typically has a considerable sample to be analyzed, which suggests the use of questionnaires with questions that are easy to be answered and that permit quantitative evaluation "a posteriori". The researcher has little or no control over the environment being observed. As with experimental research, survey research is typically applied to test models or hypotheses. An example would be a research study in which questionnaire-based data were collected from members of 300 new product development teams (i.e., teams that develop novel products such as new toothbrushes, soft drinks or car parts) in different organizations. A hypothesis to be tested in this scenario could be the degree to which a team used a particular type of technology (e.g., an electronic collaboration technology) would be correlated with the speed with which the team completed the new product development task.

### 33.2.3 Case research

This approach has its roots in general business studies, particularly those using what is frequently referred to as the “Harvard Method”. The researcher typically studies one organization, or a small sample of organizations, in depth, but usually does not participate in the organization’s day-to-day activities. Cases are analyzed either to build or validate models or theories, typically through data collection in structured and unstructured interviews. Structured interviews usually are based on a predefined set of questions; whereas unstructured interviews tend to flow more freely (a hallway conversation could be an unstructured interview). Case research is believed to be particularly useful to refute models or theories. An example would be a research study in which 3 actual business process redesign teams would be studied in-depth by a researcher, who would conduct structured interviews with the team members on a regular basis (e.g., every two weeks). The teams would be studied over a 6-month period, while they each redesigned a complex process (e.g., the process of assembling a tractor engine) in a separate organization. Each team would use a particular type of technology (e.g., an electronic collaboration technology) to a different degree. The researcher would try to understand how the technology was used by the teams, and how that use influenced the teams’ outcomes.

### 33.2.4 Action research

The main focus of this encyclopedia entry, this approach has its origins in studies of social and workplace issues. Like in case research, the researcher typically studies one organization, or a small sample of organizations, in depth. Unlike in case research, in Action Research the researcher uses participant observation and interviews as key data collection approaches. Although typically applying very little, if any, control on the environment being studied, the researcher is expected to apply some form of "positive" intervention. Typically this will be in the form of a service to the client organization. An example would be a research study in which 3 actual new product development teams would be studied in-depth by a researcher, who would also be a member of all teams. The teams would be studied over a 6-month period, while they each developed a new product in a separate organization. The researcher would be in constant contact with the other team members, which would be used in the generation of participant observation and unstructured interview notes. Each team would use a particular type of technology (e.g., an electronic collaboration technology) to a different degree, and the researcher would try to understand how the technology was used by the teams, and how that use affected the teams' outcomes.

The discussion above is somewhat artificial, since an Action Research study may employ specific techniques that are also used in experimental, survey and case research. Those three research approaches can also be conducted in unorthodox ways, borrowing techniques and ideas from other approaches. Finally, the list of approaches is not comprehensive, and does not address all possible research dimensions; for example, ethnographic research (where the researcher goes "native", becoming part of the environment being studied) may be conducted more like case research than Action Research without being either, strictly speaking. Nevertheless, hopefully the discussion above is helpful in illustrating key characteristics that differentiate Action Research from the most orthodox forms of experimental, survey and case research.

## 33.3 ACTION RESEARCH APPLIED TO HCI

No technological development has any value without action, and Action Research is all about action. Few technologies have had a stronger impact on modern society than human-computer interaction (HCI) technologies. This has been particularly true of

Internet- and Web-based HCI technologies, but is not limited to those realms. Several seminal HCI technologies predate the emergence of the Internet and the Web.

HCI emerged as a distinct area of research and practice in the early 1980s, at which point it was seen as a sub-field of or specialty area in computer science. Research on HCI has flourished worldwide, especially since the 1990s. This has been motivated by a number of factors, including the development of and experimentation with a variety of HCI tools in the 1980s and 1990s (e.g., workflow coordination and group decision support systems), the emergence of the Internet in the early 1990s, and the explosion in the personal and commercial use of the Web in the mid 1990s (motivated by the development of the first Web browsers). The flourishing of HCI research has generally coincided with the increasing use of Action Research in the study of technology-related issues.

In spite of the fact that HCI research and Action Research have grown in importance together in the last 20 years or so, there is less Action Research applied to HCI inquiry than could be expected. To be sure, there are examples of HCI studies employing Action Research, including some relatively recent ones. Nevertheless, the vast majority of the research on HCI produced in the last 15 years has employed experimental research methods, followed by survey and case research methods. Action Research trails way behind, accounting for probably no more than 5 percent of the total HCI research output.

While there is no "typical" HCI Action Research study, previous research suggests key elements that are likely to be shared by most empirical HCI studies employing Action Research, particularly studies following relaxed versions of the interpretivist and positivist epistemological paradigms. These epistemological paradigms are discussed in more detail in the next section. In these paradigms, research questions or hypotheses are formalized beforehand based on theory, and are either: (a) answered in the Action Research study (research questions); or (b) supported or refuted in the Action Research study (hypotheses). The key elements that are likely to be shared by most such empirical HCI studies employing Action Research can be summarized as follows.

### 33.3.1 Research questions or hypotheses

These are the practice-based or theory-based research questions that guide the data collection and analysis. In place of research questions, the data collection and analysis may be guided by one or more hypotheses, but this is less common in Action Research than in other research approaches (e.g., experimental research). An example of HCI research question is the following: *Does the use of a video-conferencing suite improve the quality of the outcomes generated by new product development teams whose members are geographically dispersed?*

### 33.3.2 HCI technology

This is the technology whose impact on a research client is the main subject of the research. An example of HCI technology is a video-conferencing suite. More than one HCI technology may be studied in an HCI Action Research study.

### 33.3.3 Practical problems

These are the problems being faced by an individual, group or organization; which the HCI Action Research study aims at solving, at least in part. Some prefer to refer to practical problems by using a more “benign” term, namely that of “opportunities for improvement”. An example of practical problem is the following: *New products need to be constantly developed by geographically dispersed teams, but the transportation and lodging costs associated with bringing team members together currently prevent more than two thirds of the needed teams from being conducted.*

### 33.3.4 Research client

This is the individual, group, or organization whose practical problem (or problems) is supposed to be solved by the HCI Action Research study. An example of a research client would be an automobile manufacturer with several factories in Europe, the U.S.A. and South America.

One of the most straightforward and efficient ways of conducting an HCI Action Research study is to collect data using the same instrument (e.g., a questionnaire) at two key points in time, namely before and after the introduction of the HCI technology, and add to that other types of data collection such as participant observation and unstructured interview notes.

The technology introduction would more often than not have the goal of solving an important practical problem being faced by the research client. Usually, it is a good idea to collect quantitative as well as qualitative data before and after the technology introduction. The quantitative data can be used in simple non-parametric comparison of means analyses, whereas the qualitative data can be used to find explanations and underlying causes for the patterns observed in the data.

In spite of its simplicity, the type of research design discussed above is relatively rare in Action Research. It is much more common to see published examples of Action Research in which only qualitative data is collected, mostly during and after the Action Research intervention (e.g., HCI technology introduction). This may be a problem, as the researcher may end up “drowning” in a “sea of data” from which multiple models can be derived. Research project planning may also be hampered; research usually is conducted on a fixed budget and within a limited timeframe.

Quite often Action Research studies are conducted through multiple iterations of what has become known as the “Action Research cycle”, rather than a “one shot,” non-cyclical research design. That is, quite often Action Research studies are longitudinal, as opposed to cross-sectional, although this is not always the case. Cyclic Action Research is, generally speaking, a good thing. The Action Research cycle involves the identification of practical problems, the solution of those problems, and reflection on the part of the researcher, which is then followed again by the identification and solution of problems, new reflection, and so on. Conducting multiple iterations of the Action Research cycle tends to add validity and credibility to the research findings, as repeated observations in various iterations lead to the identification of clear patterns.

The most widely referenced version of the "Action Research cycle" has been proposed by Gerald Susman and Roger Evered, in 1978 (see Figure 33.1). It comprises five stages: diagnosing, action planning, action taking, evaluating, and specifying learning. The diagnosing stage, where the cycle begins, involves the identification of an improvement opportunity or a general problem to be solved at the client organization. The following stage, action planning, involves the consideration of alternative courses of action to attain the improvement or solve the problem identified. The action taking stage involves the selection and implementation of one of the courses of action considered in the previous stage. The evaluating stage involves the study of the outcomes of the selected course of action. Finally, the specifying learning stage involves reviewing the outcomes of the evaluating stage and, based on this, knowledge building in the form of a model describing the situation under study. In studies that involve several iterations of the Action Research cycle, the specifying learning stage is followed by the diagnosing stage of a subsequent cycle, which can take place in the same organizational context or in a different one (e.g., a different department or company).

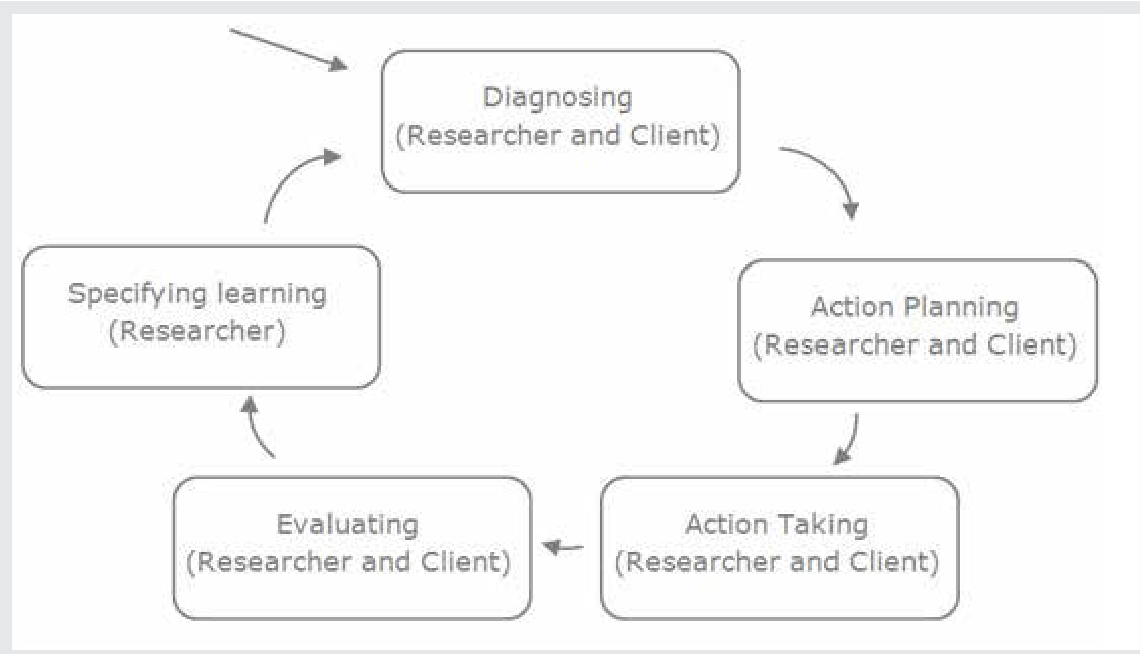



FIGURE 33.1: Susman and Evered's (1978) Action Research cycle.



## 33.4 ACTION RESEARCH AND EPISTEMOLOGIES

Epistemologies can be seen as systems of concepts, rules, and criteria that find acceptance among a community of researchers as a basis for the generation of what that community of researchers sees as valid knowledge. By far the most widely subscribed epistemology among empirical HCI researchers is positivism.

Research that conforms to positivist inquiry tenets usually departs from a set of theoretical propositions or hypotheses, and aims at testing those propositions or hypotheses through the analysis of empirical data. Also, in positivist research the data is usually (although not always) of a quantitative nature. The research methods employed in positivist studies often reflect those traditionally used by natural scientists.

Experimental research is usually conducted in a positivist way. Therefore, our previous example of experimental research can be invoked to illustrate typical positivist research: a research study in which a group of people used a technology (e.g., an online collaborative writing tool) to perform a group-based task (e.g., collaborative writing of a sales contract), and another group performed the task by interacting face-to-face. Again, an example of hypothesis could be that the technology group would write a "better" contract than the no-technology group.

In the interpretivist (a.k.a. interpretive) epistemological tradition the investigator usually departs from research questions to be answered through data collection and analysis, not hypotheses. Interpretivism is frequently referred to as anti-positivism, in an epistemological sense. Often theoretical models are built based on the analysis of qualitative data, and deep reflection by the researcher on the study subjects' and the researcher's own subjective views of the world. The critical epistemology is more similar to the interpretivist than to the positivist epistemology, but here the investigator typically wants to empower the subjects of the research so that they can correct a situation that is perceived as unfair and/or oppressive.

One issue that has led to some debate among Action Research scholars in the past is whether Action Research can be conducted in ways that are consistent with different epistemologies, including the positivist epistemology. The debate has been motivated by the fact that Action Research has typically been used in research studies that do not conform very well with traditional positivist standards, and that are better aligned with what many would see as the interpretivist and critical epistemologies. In fact, one could argue that today there is resistance in scholarly Action Research circles against the notion of positivist Action Research, and that resistance can be quite strong within specific Action Research communities. Examples are the Action Research communities that conduct investigations according to two main Action Research frameworks: John Heron's cooperative inquiry, and Paulo Freire's participatory action research.

The above scenario creates a problematic situation û what one could reasonably call a vicious circle. Empirical HCI research, where the behavioral impact of HCI technologies is evaluated, has traditionally been and continues being overwhelmingly positivist in nature. There are practical reasons for this status quo û positivist research is easier to plan and conduct, and the quantitative form of the results is usually very clear-cut.

However, because of this positivist empirical HCI research tradition, researchers who try to employ Action Research to study HCI are hampered not only once but twice in their efforts. On one hand, they have to justify using Action Research in a positivist manner, which is likely to meet with opposition from Action Research scholars. On the other hand, they have to sell the notion that Action Research can be useful for HCI research, which is likely to be seen with suspicion by established HCI researchers, who often view Action Research is a “soft” investigative approach.

This is an unfortunate state of affairs, because Action Research can address a key problem with past HCI research, namely its lack of “real world appeal”. In other words, since past empirical HCI research has been by and large based on laboratory experiments with students, it has been difficult for practicing managers and professionals to personally relate to and benefit from many of the findings resulting from that research. Those managers frequently see research conducted in controlled laboratory settings as leading to findings that carry little external validity.

Can Action Research be successfully employed in empirical HCI research? The answer to this question is certainly “yes”, and there are several examples of that (see, e.g., Kock, 1998; Kock et al., 2000; Kock and DeLuca, 2007). Can it be done in a positivist way? Well, based on some various examples, the answer to this follow-up question also seems to be “yes” (see, e.g., Davis, 2001; DeLuca et al., 2006). The key here is perhaps to be creative so that certain characteristics of Action Research are used to add strengths to more traditional HCI inquiry.

A widely overlooked strength of Action Research comes from the observation that it exposes the researcher to significantly more data (although relatively sparse) than more focused research approaches (e.g., experimental and survey research). One could adopt Karl Popper's view that exposure to a large body of data, whose analysis does not uncover evidence that contradicts a hypothesis, is in fact "evidence" in support of the hypothesis. Given this, Action Research could be seen as quite adequate for positivist HCI inquiry. This is a modified positivist stance, which has been called post-positivist, and sometimes neo-positivist.

## 33.5 FUTURE DIRECTIONS

The applied nature of Action Research has tremendous appeal. At the same time, it usually makes Action Research much more difficult to implement in practice. This creates difficulties for the use of Action Research in academic circles. A lot of research is done in the context of doctoral dissertations. And a good doctoral dissertation is a finished doctoral dissertation. Adopting a research approach that makes research more difficult to complete is not the best way to get a doctoral degree.

So is Action Research doomed then? Not really, due to two interesting phenomena: (a) increasingly, research is being conducted with funding from outside academia; and (b) more and more research is being done outside academia.

External funding agencies often value practical research. The reason is that their constituents are particularly interested in research whose results have very practical applications, be they organizational or community stakeholders. While this is especially true of non-government organizations, it also applies to government agencies. Examples are the National Science Foundation in the U.S.A. and the European Commission in Europe. Both have recently ramped up research funding for HCI-related studies. The European Commission tends to favor Action Research-like research more than the National Science Foundation, but the latter is catching up quickly (Kock and Antunes, 2007).

Outside academia, research practical applications is becoming widespread. Companies like Google and Microsoft employ a significant proportion of their workforce in research and development divisions. Many of those researchers focus on Action Research-like HCI research. Notable pioneers in this respect are Jonathan Grudin at Microsoft, and Craig Nevill-Manning at Google. The careers of these two technology professionals and researchers, whom I have been following, provide a glimpse of the future of Action Research-like research in HCI. I have had many discussions with Jonathan about HCI issues, especially the connection with Darwinian evolution. Craig and I were doctoral students at about the same time at the University of Waikato, in New Zealand.

## 33.6 WHERE TO LEARN MORE

### 33.6.1 Special issues of journals

HCI researchers often identify themselves with broader research communities. One such community is that of information systems researchers. With that in mind, a couple of special issues on information systems Action Research are worth checking, as they provide exemplars of Action Research studies that can be used as a basis for HCI researchers interested in employing Action Research. The first is the special issue on Action Research in information systems published in the journal Information Technology & People in 2001 (volume 14, number 1). The second is the special issue on Action Research in information systems published in the journal MIS Quarterly in 2004 (volume 28, number 3).

The special issue published in the journal *Information Technology & People* in 2001 was the first special issue ever on Action Research in information systems. The issue contained six articles. Three of those are conceptual, in the sense that they are aimed at providing insights on how to conduct information systems Action Research. The other three articles are empirical, in the sense that they discuss actual information systems Action Research studies and their results. Of the empirical articles, two addressed HCI issues in the context of group support systems investigations. Those articles are: GSS and Action Research in the Hong Kong Police by Robert Davison; and Action Learning and Groupware Technologies: A Case Study in GSS Facilitation Research, by Pak Yoong, Brent Gallupe.

The special issue on Action Research in information systems published in the journal *MIS Quarterly* in 2004 was aimed at providing a set of exemplars of information systems Action Research studies of an empirical nature. As such, all of the six articles published in this special issue report on empirical studies that employed Action Research to investigate information systems phenomena. None of the articles seems to be aimed at squarely addressing HCI issues, although at least two of the articles address issues that are likely to be directly relevant for HCI researchers. Those articles are: Informating the Clan: Controlling Physicians Costs and Outcomes, by Rajiv Kohli and William J. Kettinger; and Small Business Growth and Internal Transparency: The Role of Information Systems, by Christopher T. Street and Darren B. Meister.

### 33.6.2 Books and research articles

*Davis, Erica (2001): Applying a Personalized System of Instruction to Internet-Based Training. Doctoral Dissertation. Philadelphia, USA, Temple University*

*DeLuca, Dorothea, Gasson, Susan and Kock, Ned (2006): Adaptations that Virtual Teams Make so that Complex Tasks Can be Performed Using Simple e-Collaboration Technologies. In International Journal of e-Collaboration, 2 (3) pp. 64-90*

*Kock, Ned (ed.) (2006): Information Systems Action Research: An Applied View of Emerging Concepts and Methods. New York, USA, Springer*

*Kock, Ned (2004): The Three Threats of Action Research: A Discussion of Methodological Antidotes in the Context of an Information Systems Study. In Decision Support Systems, 37 (2) pp. 265-286*

*Kock, Ned (1998): Can Communication Medium Limitations Foster Better Group Outcomes? An Action Research Study. In Information and Management, 34 (5) pp. 295-305*

*Kock, Ned (ed.) (2008): Darwinian Perspectives on Electronic Communication: Special Issue of the Journal IEEE Transactions on Professional Communication. IEEE Computer Society Press*

Kock, Ned (ed.) (2008): *Encyclopedia of E-Collaboration. Information Science Reference*

Kock, Ned and Antunes, Pedro (2007): *Government Funding of E-Collaboration Research in the European Union: A Comparison with the United States Model. In International Journal of e-Collaboration, 3 (2) pp. 36-47*

Kock, Ned and DeLuca, Dorothea (2007): *Improving Business Processes Electronically: An Action Research Study in New Zealand and the US. In Journal of Global Information Technology Management, 10 (3) pp. 6-27*

Kock, Ned, Auspitz, Camile and King, Brad (2000): *Using the Web to Enable Industry-University Collaboration: An Action Research Study of a Course Partnership. In Informing Science, 3 (3) pp. 157-167*

Kohli, Rajiv and Kettinger, William J. (): *Informating the Clan: Controlling Physicians Costs and Outcomes. In MIS Quarterly, 28 (3)*

Street, Christopher T. and Meister, Darren B. (2004): *Small Business Growth and Internal Transparency: The Role of Information Systems. In MIS Quarterly, 28 (3)*

Susman, Gerald I. and Evered, Roger D. (1978): *An Assessment of the Scientific Merits of Action Research. In Administrative Science Quarterly, 23 (4) pp. 582-603*

### 33.6.3 Online Resources

Action Learning and Action Research Association

City University of New York's Participatory Action Research and Design Collective

Himalayan Action Research Centre

Illinois Wesleyan University's Action Research Center

Institute for Community Research's Youth Action Research Institute

Pepperdine University's Center for Collaborative Action Research

Southern Cross University's Action Research Resources

University of Bath's Center for Action Research in Professional Practice

The Paulo and Nita Freire International Project for Critical Pedagogy

The Tavistock Institute of Human Relations

University of Cincinatti's Action Research Center

Wikipedia's Article on Action Research

Wikipedia's Article on Kurt Lewin

Women's Health and Action Research Centre

## 33.7 APPENDIX: DOCTORAL ACTION RESEARCH ON HCI

A great deal of the research output produced every year, and published in academic journals, is the direct result of doctoral research investigations. The field of HCI is no exception to this general rule, so it is a good idea to contemplate the pros and cons of conducting doctoral research on HCI issues employing Action Research.

One of the best ways to get a doctoral degree is to test an existing theory and, based on the results, add to or refine the theory. It is not very wise to try to develop a new theory as part of one's doctoral research project. Many doctoral students are prone to think of their research projects as likely to lead to theoretical insights that will change the world in a major way. Nevertheless, it is unlikely that doctoral students' ideas will have the same impact as Darwin's theory of evolution, or Einstein's theory of relativity; which were not developed as part of Darwin's or Einstein's doctoral work, by the way.

Conducting research aimed at testing an existing theory is quite likely to lead someone's research to fall into the general epistemological category called positivist research, discussed earlier. And, as previously argued, there is nothing wrong with conducting Action Research in a positivist manner. However, one problem may arise. Traditionally, Action Research has not been seen as the best approach for the conduct of positivist inquiry. In fact, Action Research has been widely viewed as an ideal approach to create new theories grounded in action-oriented projects, particularly in organizational settings.

So, what is a doctoral student to do when contemplating using Action Research to investigate HCI issues? First, it would be advisable to have a look at recent examples of doctoral dissertations that accomplished this. Second, it is highly advisable to design the research in a positivist (or post-positivist) manner, following some of the suggestions provided earlier. Finally, the student should make sure that the doctoral dissertation committee members are receptive to the idea of Action Research being conducted in a positivist manner. After all, those committee members are ultimately the ones that will decide whether the degree is granted or not. Those who employ and/or subscribe to the Action Research approach known as “canonical Action Research” are likely to be so inclined, and others who are not can be educated based on publications discussing what is known as “canonical Action Research”.

Nevertheless, a number of obstacles await those doctoral students who decide to employ Action Research to study HCI issues. Those students who opt for studying HCI effects in organizational settings, for example, will face the challenge of finding one or more organizations willing to work with them. Even when organizational support is achieved, there is the danger that the support will be withdrawn before enough research data is collected. Finally, a multitude of political issues will have to be dealt with. There may be suspicion and opposition by employees, if support is obtained from the organization’s management first, without much grass-roots consultation. Dealing with such political issues is likely to ensure that the doctoral student employing Action Research will have to spend significantly more time and effort with the research project than doctoral students employing more traditional HCI research approaches (e.g., experimental research).

Yet, the enhanced credibility of the research findings, the excitement of being “part of the action”, and the personal satisfaction that comes from helping improve people’s lives while conducting an Action Research study, may be well worth all the extra effort. This is why I conducted my doctoral research, many years ago, using Action Research to address HCI issues.

## 33.8 COMMENTARY BY SUSANNE BØDKER

How to cite this commentary in your report

### Susanne Bødker

© *Susanne Bødker*

Susanne Bødker is professor of Human Computer Interaction at the Computer Science Department, University of Aarhus. Her research areas include participatory design, computer-supported cooperative work and human-computer interaction. Her PhD thesis, Through the Interface – a Human Activity Approach to User Interface Design was an early attempt to present activity theoretical HCI to an int...

Susanne Bødker
Susanne Bødker is a member of The Interaction Design Foundation

Let me start this comment by saying that I'm in total agreement with the comments made by Bjørn Erik Munkvold in his commentary (see below). Instead of repeating his arguments I will dig a little deeper into some details in the arguments and history outlined by Kock.

Back in the beginning of the 1990s Bannon (1991) identified the move from first wave, cognitivist HCI to the second wave, which was theoretically more diverse, focusing more on the active role of users. In his writing about the move "from human factors to human actors" he pointed both towards a different way of thinking about users, as active human beings, and towards the roles that these users may play in design and research.

In the second wave, focus was on groups working with a collection of applications. Theory focused on work settings and interaction within well-established communities of practice. Rigid guidelines, formal methods, and systematic testing were mostly abandoned for proactive methods such as a variety of participatory design workshops, prototyping and contextual inquiries (Bødker 2008). With this picture in mind, and at this quite general level, I have difficulties recognizing the picture of the past 20 years of HCI painted by Kock: "the vast majority of the research on HCI produced in the last 15 years has employed experimental research methods followed by survey and case research methods." Part of my problem in this is perhaps my own lack of ease of making such clear distinctions.

As pointed out in Bjørn Erik Munkvold's commentary, an important source of inspiration for the second wave of HCI was a number of Scandinavian projects, applying participatory design with future users of technology (in the workplace) and action research with systems developers to improve their ways of working. Without going into the intrinsic details about the genealogy of these projects and their methodological roots, it feels safe to say that they were in turn inspired by Nordic work researchers like Thomas Mathiesen and Bjørn Gustavsen, who in turn were inspired by Kurt Lewin. When looking at Kock's historical line-up, however, these projects approached action research largely by taking an active stance against a positivistic view. This was in parts a result of significant influence from Marxist thinking, but also of influence by activists like Paulo Freire, who gets mentioned by Kock.

In Finland a parallel development happened leading to Engeström's focus on work development research (1987), where Lewin's thinking has been important together with activity theory. This development had further strong parallels in Germany. Not being a researcher in the history of ideas I will leave this discussion, except for saying that in my view it has been through this lineage that action research has had the strongest impact on HCI, in all modesty e.g. through my own work.

Kock's characteristics of the current state of affairs in HCI does not match with Bannon's characteristics of the second wave of HCI, and with the current development into the third wave (Bødker 2006, Harrison et al 2007). As a matter of fact, lab studies have played a very insignificant role in conferences like ACM CHI or Interact over the past 20 years, and it is not many lab papers I see in my everyday life as editor of ACM ToCHI and IJHCS. Given that, Kock may well be right that what gets applied are experimental research methods followed by survey and case research methods, but they are applied in much more mixed manners than suggested by Kock. This is largely due to the quite significant impact of qualitative, interpretative, and even radical humanist methods (Burell & Morgan 1979) in second wave HCI (e.g. ethnomethodology through Suchman (1987), activity theory through Kaptelinin, Nardi, Kuutti and Bødker). Actually one could probably argue that there is too little systematic research and too little theory in contemporary HCI in general, and with this statement I strongly sympathize with Kock's attempt to more rigorously outline the methods and outcomes of different research approaches in HCI.

I am however confused by Kock's comment about HCI's current lack of real-world appeal and its accompanying difficulties for the impact of action research. Having just come back from the DIS 2010 conference where Yvonne Rogers talked about HCI's "turn to the wild" while numerous researchers talked about the role of design in HCI research, I am convinced that Kock is wrong in this analysis:

At a time when use contexts and application types are broadened, and intermixed across private and public spheres; where technology spreads from the workplace to our homes and everyday lives and culture; where new elements of human life are included in the human-computer interaction such as culture, emotion and experience, there is more than ever a need for action-based approaches and interventions in HCI. But more than ever, these need to take an active stance against positivistic thinking.

## References

Bannon, Liam. (1991) From Human Factors to Human Actors: The role of psychology and human-computer interaction studies in systems design. In: Greenbaum, Joan & Kyng, Morten. (Eds.) (1991) Design at Work. Cooperative Design of Computer Systems. Lawrence Erlbaum Associates, pp. 25-44.

Burrell, G., & Morgan, G. Sociological Paradigms and Organizational Analysis, Heinemann, 1979, 1-37

Bødker, Susanne. (1991). Through the Interface. A Human Activity Approach to User Interface Design. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Bødker, Susanne. (2006). When second wave HCI meets third wave challenges. ACM International Conference Proceeding Series; Vol. 189. Proceedings of the 4th NordiCHI, ACM, pp. 1 – 8.

Engeström, Y. (1987). Learning by expansion. Helsinki: Orienta Konsultit.

Steve Harrison, Deborah Tatar, and Phoebe Sengers. "The Three Paradigms of HCI." alt.chi 2007.

Kaptelinin, Victor. & Nardi, Bonnie. (2006) Acting with technology, Cambridge, MA: MIT press.

Kuutti, K., The Concept of Activity as a Basic Unit for CSCW Research., in Proceedings of the 2nd ECSCW., L.J. Bannon, M. Robinson, and K. Schmidt, Editors. 1991, Kluwer: Amsterdam. p. 249-264.

Munkvold, Bjørn Erik [commentary]

Suchman, Lucy. (1987). Plans and Situated Actions, Cambridge UK: Cambridge University Press.

## 33.9 COMMENTARY BY BJØRN ERIK MUNKVOLD

How to cite this commentary in your report

### Bjørn Erik Munkvold



Bjørn Erik Munkvold is Professor of Information Systems and Director of the PhD Programme in Information Systems at University of Agder, Norway. His main research interests include e-Collaboration and virtual work, organizational IT implementation and qualitative research methododology...

Bjørn Erik Munkvold
Bjørn Erik Munkvold is a member of The Interaction Design Foundation

Ned Kock's chapter offers an insightful introduction to the concept of Action Research and how this could be related to the field of HCI. The chapter reflects Ned's broad experience with this type of research, and will be useful to anyone looking for an overview of the characteristics and challenges of applying this research approach in the context of HCI. The chapter also points to sources for learning more about the topic. In this brief commentary I will address some issues from the chapter that I think could be elaborated more, to bring out the key characteristics of Action Research as well as some nuances in our understanding of this.

Ned's point of departure for the chapter is that to understand Action Research requires being able to differentiate this from other types of research. As presented in the chapter, a focus on *action* and *practice* are key elements of action research projects. However, by themselves these elements are necessary but not sufficient foci for a study to be classified as action research. For example, much of the research on IT-enabled change in organizations (conducted as case studies or surveys) focuses on changes to practice through technology-related actions. Yet, this research will typically not be classified as action research, as long as the researcher role is merely that of an observer. Thus, the role of the researcher(s) is a key distinguishing element in Action Research, in that an 'action researcher' would normally be closely involved in specifying and bringing about the targeted change. This is done in close collaboration with the 'client', bringing me to the next point of the collaborative nature of Action Research. As specified in the Action Research cycle of Susman and Evered (1978) in the chapter, the steps in the process are conducted in collaboration between the research and the client. Even in the final stage of specifying learning, while the researcher may be responsible also this step would normally involve mutual learning between the researcher and clients/practitioners. The focus of joint collaboration in the research process is thus a key distinguishing feature of Action Research, emphasized in the early literature defining this research (e.g. Rapoport, 1970; Susman and Evered, 1978). This does not go against what is presented in the chapter, but intends to emphasize more the important role of the researcher as 'change agent' and the collaborative model of action research.

Related to the application of action research in HCI, the chapter raises an issue of a possible mismatch between the positivist epistemology of HCI and the non-positivist nature of action research. The argument made is that since action research could also be done in a positivist manner, it should be 'acceptable' for the HCI community. My view on this issue is somewhat different. First, as discussed by Carol (2009), the HCI community today is very broad and from my European/Scandinavian perspective I do not experience a similar strong dominance of the positivist perspective. The part of the HCI community focusing on participatory design here serves as an example. Second, I belong to those who regard the nature of Action Research to be based on an interpretivist perspective, finding it difficult to reconcile the engaged role of the action researcher with a positivist epistemology. This of course does not preclude combining qualitative and quantitative methods, as advocated in the chapter. Thus, rather than giving Action Research a positivist framing, I would argue that this research approach should be made attractive to the HCI community on its merits of increased relevance through building and evaluating IT artifacts in close interaction with practice.

An increasing focus on action research could also be seen as a response to Van de Ven's (2007) call for engaged scholarship, defined as "a participative form of research for obtaining the different perspectives of key stakeholders (researchers, users, clients, sponsors, and practitioners) in studying complex problems" (p. 9). Action Research has also been focused lately as part of the increasing interest in design science (Iivary, 2007), and the debate on similarities and differences between these two approaches. A recent contribution is the Action Design Research method suggested by Sein et al. (2010), that combines the design science focus of building innovative IT artifacts with the Action Research focus of learning from the intervention when applying the artifact for solving a problem in practice.

The chapter appendix addresses an important issue about the challenges of conducting doctoral research based on Action Research. The potential obstacles discussed resonate well with experiences discussed by Jesper Simonsen from his action research projects in the Scandinavian context (Simonsen, 2009). He suggests that some of these challenges of Action Research can be mitigated by only assigning the PhD student with specific parts of the Action Research project, and having the supervisor and fellow senior researchers co-participate in the project. In any case, Action Research projects will often be undertaken by a research team rather than a single researcher.

As pointed out in the chapter, Action Research offers a great potential for HCI research. This is exemplified by the strong tradition of Action Research in the Scandinavian IS community over the years (Mathiassen and Nielsen, 2008). In addition to improving business practices through IT, Action Research also holds promise as a basis for research on how to develop and apply technology to bring about positive change in important societal areas such as healthcare and environment.

## References

Carroll, John M. (2009). *Encyclopedia entry on Human Computer Interaction (HCI)*. Retrieved 13 July 2010 from Interaction-Design.org: http://www.interaction-design.org/encyclopedia/human_computer_interaction_hci.html

Iivari, Juhani (2007): *A paradigmatic analysis of information systems as a design science*. In Scandinavian Journal of Information Systems, 19 (2), pp. 39-64.

Mathiassen, Lars and Nielsen, Peter A. (2008): Engaged Scholarship in IS Research: The Scandinavian Case. In Scandinavian Journal of Information Systems, 20 (2), pp. 3-20.

Rapoport, Robert N. (1970): *Three dilemmas of action research*. Human Relations, 23, pp. 499-513.

Sein, Maung K., Henfridsson, Ola, Purao, Sandeep, Rossi, Matti and Lindgren, Rikard (2010): *Action Design Research*. Forthcoming in MIS Quarterly

Simonsen, Jesper (2010): A Concern for Engaged Scholarship: The Challenges for Action Research Projects. In Scandinavian Journal of Information Systems, 21 (2), pp. 111-128.

Susman, Gerald I and Evered, Roger D. (1978): *An assessment of the scientific merits of action research*. Administrative Sciences Quarterly, 23, pp. 582-603.

Van de Ven, Andrew H. (2007): *Engaged Scholarship: A Guide for Organizational and Social Research*, Oxford, Oxford University Press.

## 33.10 COMMENTARY BY RICHARD BASKERVILLE

How to cite this commentary in your report

### Richard Baskerville

*© Richard Baskerville*

Baskerville is a Professor in the Department of Computer Information Systems, J. Mack Robinson College of Business at Georgia State University. His research specializes in security of information systems, methods of information systems design and development, and the interaction of information systems and organizations. His interests in methods extend to qualitative research methods. B...

Richard Baskerville
Richard Baskerville is a member of The Interaction Design Foundation

Ned Kock's article, "Action Research: Its Nature and Relationship to Human-Computer Interaction", which appears in the HCI Encyclopedia of Interaction-Design.org, provides an excellent and stimulating introduction and overview into Action Research for the HCI researcher. There is little with which experts may disagree, except over the lacunae that are normally driven by space limitations.

Our moniker includes the word "research" and this part of the term leads scholars to regard the technique in the context of other research approaches, as Kock does in his comparison with experiments, surveys, etc. But action research's origins are more humble. It originated primarily as a means for researching a practical problem, usually something along the lines of, "How do we get our organizational members to stop behaving badly?" Action research is not only for academic research or even for research-and-development. It is also a means by which an organization can undertake self-therapy to make itself better. Such non-academic action research may not spill out top publications, but it can cure an organization's ills. Little wonder it is a useful approach for consultants seeking to help those in need of organizational development (Baskerville & Wood-Harper, 1998).

In one sense, action research is such a painfully simple idea that the many layers of epistemology, methodology, and infrastructure might seem superfluous. Why bother? We bother because this rigor is necessary to those researchers (such as the academics) that plan to capture the general knowledge gained in the action research and introduce this into the scientific literature. It is made necessary if we are to produce more than just the practical solution, but also the "credentialed" knowledge that can be proven valid and reliable. Kock is on-target with his pluralist discussion of action research epistemology. After all, philosophy is not an attribute of the research method; it is an attribute of the philosopher who uses it (Baskerville, 1991).

In his appendix on Doctoral Action Research, Kock introduces us to some of the issues confronting student use of the approach. For me, these issues spring from the limited control that an action researcher can exercise over his/her own research process. Most action research, and especially HCI action research, is sociological in nature. When the context is social, "control" is problematic. Most commonly, action research projects unfold under the shared control by researchers and subjects. Under such shared decision-making, theories that guide the project may change, plans for action may change, and participants may change, all beyond the unilateral control of the researcher. Doctoral researchers (and their directors) have to understand that action research is often problem-centric rather than theory- or question-centric (Avison, Baskerville, & Myers, 2001).

For information systems, action research is increasingly contextualized within Van de Ven's (2007) notion of "Engaged Scholarship" (along with design science), as noted in Munkvold's commentary on Kock's article (in this volume). The original conceptualization of engaged research was as a fifth form of scholarship beyond four other forms: discovery, teaching, application, and integration (Boyer, 1996). In other words, because engaged research is different from discovery research, it is seen as useful only in its humblest sense; as a means for practical problem solving. Van de Ven's work is important because it integrates the scholarship of discovery with engaged scholarship. This integration fits well with the use of action research by scholars.

HCI is often concerned with design. Action research is sometimes mistakenly conflated with design science research (Järvinen, 2007). After all, "design" (as a verb) is "action". But the paradigms underlying these two research approaches are quite different (Iivari, 2007). While action research is grounded in social psychology (hence centering the human), design science research is grounded in engineering economics (hence centering the artifact, Simon, 1996). For HCI researchers, it is important to recognize when the two approaches are used in isolation, and when these are being integrated. When integrated as a process, the underlying epistemology has to be integrated with the care that Kock's work suggests.

## References

Avison, D., Baskerville, R., & Myers, M. (2001). Controlling action research projects. *Information Technology and People, 14*(1), 28-45.

Baskerville, R. (1991). Practitioner Autonomy and the Bias of Methods and Tools. In H.-E. Nissen, H. K. Klein & R. A. Hirschheim (Eds.), *Information Sys-*

*tems research: Contemporary Approaches & Emergent Traditions* (pp. 673-698). Amsterdam: North-Holland.

Baskerville, R., & Wood-Harper, A. T. (1998). Diversity in Information Systems Action Research Methods. *European Journal of Information Systems, 7*(2), 90-107.

Boyer, E. H. (1996). The scholarship of engagement. *Journal of Public Service and Outreach, 1*, 11-20.

Iivari, J. (2007). A paradigmatic analysis of Information Systems as a design science. *Scandinavian Journal of Information Systems, 19*(2), 39-63.

Järvinen, P. (2007). Action Research is Similar to Design Science *Quality and Quantity 41*(1), 37-54.

Simon, H. A. (1996). *The Sciences of the Artificial* (3rd ed.). Cambridge, Mass.: MIT Press.

Van de Ven, A. H. (2007). *Engaged Scholarship: A Guide for Organizational and Social Research*. Oxford: Oxford University Press.

## 33.11 REFERENCES

*Carroll, John M. (2012). Human Computer Interaction - brief intro. Retrieved 6 June 2013 from [URL to be defined - in press]*

*Carroll, John M. (ed.) (2003): HCI Models, Theories, and Frameworks. San Francisco, Morgan Kaufman Publishers*

*Davis, Erica (2001): Applying a Personalized System of Instruction to Internet-Based Training. Doctoral Dissertation. Philadelphia, USA, Temple University*

*DeLuca, Dorothea, Gasson, Susan and Kock, Ned (2006): Adaptations that Virtual Teams Make so that Complex Tasks Can be Performed Using Simple e-Collaboration Technologies. In International Journal of e-Collaboration, 2 (3) pp. 64-90*

*Kock, Ned (2004): The Three Threats of Action Research: A Discussion of Methodological Antidotes in the Context of an Information Systems Study. In Decision Support Systems, 37 (2) pp. 265-286*

*Kock, Ned (1998): Can Communication Medium Limitations Foster Better Group Outcomes? An Action Research Study. In Information and Management, 34 (5) pp. 295-305*

*Kock, Ned (ed.) (2008b): Darwinian Perspectives on Electronic Communication: Special Issue of the Journal IEEE Transactions on Professional Communication. IEEE Computer Society Press*

*Kock, Ned (ed.) (2006): Information Systems Action Research: An Applied View of Emerging Concepts and Methods. New York, USA, Springer*

*Kock, Ned (ed.) (2008a): Encyclopedia of E-Collaboration. Information Science Reference*

*Kock, Ned and Antunes, Pedro (2007): Government Funding of E-Collaboration Research in the European Union: A Comparison with the United States Model. In International Journal of e-Collaboration, 3 (2) pp. 36-47*

*Kock, Ned, Auspitz, Camile and King, Brad (2000): Using the Web to Enable Industry-University Collaboration: An Action Research Study of a Course Partnership. In Informing Science, 3 (3) pp. 157-167*

## About the author

### Ned Kock

Has also published under the name of:
"Nereu F. Kock, Nereu Kock"

Personal Homepage:
http://www.nedkock.com

Current place of employment:
Texas A&M International University

**Ned Kock** is Professor of Information Systems and Director of the Collaborative for International Technology Studies at Texas A&M International University. He holds degrees in electronics engineering (B.E.E.), computer science (M.S.), and management information systems (Ph.D.). Ned has authored and edited several books, including the bestselling Sage Publications book titled *Systems Analysis and Design Fundamentals: A Business Process Redesign Approach*. He has published his research in a number of high-impact journals including *Communications of the ACM*, *Decision Support Systems*, *European Journal of Information Systems*, *European Journal of Operational Research*, *IEEE Transactions (various)*, *Information & Management*, *Information Systems Journal*, *Journal of the Association for Information Systems*, *MIS Quarterly*, and *Organization Science*. He is the Founding Editor-in-Chief of the *International Journal of e-Collaboration*, Associate Editor for Information Systems of the journal *IEEE Transactions on Professional Communication*, and Associate Editor of the *Journal of Systems and Information Technology*. His main research interests are biological and cultural influences on human-technology interaction, non-linear structural equation modeling, electronic communication and collaboration, action research, ethical and legal issues in technology research and management, and business process improvement.

## YOUR NOTES AND THOUGHTS ON CHAPTER 33

Record your notes and thoughts on this chapter. If you want to share these thoughts with others online, go to the bottom of the page at:
http://www.interaction-design.org/encyclopedia/action_research.html

**NOTES:**

# Chapter 35

# Data Visualization for Human Perception

*by Stephen Few.*

Data visualization is the graphical display of abstract information for two purposes: sense-making (also called data analysis) and communication. Important stories live in our data and data visualization is a powerful means to discover and understand these stories, and then to present them to others. The information is abstract in that it describes things that are not physical. Statistical information is abstract. Whether it concerns sales, incidences of disease, athletic performance, or anything else, even though it doesn't pertain to the physical world, we can still display it visually, but to do this we must find a way to give form to that which has none. This translation of the abstract into physical attributes of vision (length, position, size, shape, and color, to name a few) can only succeed if we understand a bit about visual perception and cognition. In other words, to visualize data effectively, we must follow design principles that are derived from an understanding of human perception.

As the saying goes, “a picture is worth a thousand words” - often more - but only when the story is best told graphically rather than verbally and the picture is well designed. You could stare at a table of numbers all day and never see what would be immediately obvious when looking at a good picture of those same numbers. Allow me to illustrate. Here’s a simple table of sales data - a year’s worth - divided into two regions:

2009 Sales (thousands of U.S. $)

| Region | Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct | Nov | Dec | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Domestic | 1,983 | 2,343 | 2,593 | 2,283 | 2,574 | 2,838 | 2,382 | 2,634 | 2,938 | 2,739 | 2,983 | 3,493 | 31,783 |
| International | 574 | 636 | 673 | 593 | 644 | 679 | 593 | 139 | 599 | 583 | 602 | 690 | 7,005 |
| Total | 2,557 | 2,979 | 3,266 | 2,876 | 3,218 | 3,517 | 2,975 | 2,773 | 3,537 | 3,322 | 3,585 | 4,183 | 38,788 |

FIGURE 35.1



This table does two things extremely well: it expresses these sales values precisely and it provides an efficient means to look up values for a particular region and month. But if we’re looking for patterns, trends, or exceptions among these values, if we want a quick sense of the story contained in these numbers, or we need to compare whole sets of numbers rather than just two at a time, this table fails.

Now look at the following picture of the same information in the form of a line graph:

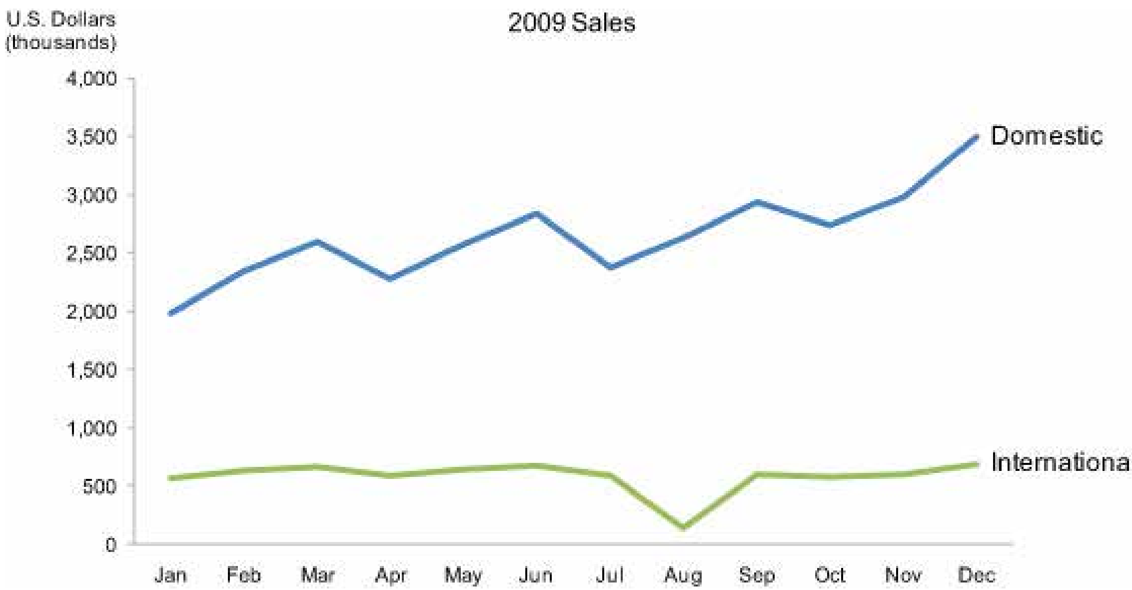



Figure 35.2

*Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.*

Several facts now leap into view:

- Domestic sales were considerably and consistently higher than international.
- Domestic sales trended upward over the year as a whole.
- International sales, in contrast, remained relatively flat, with one glaring exception: they decreased sharply in August.

Domestic sales exhibited a cyclical pattern - up, up, down - that repeated itself on a quarterly basis, always reaching the peak in the last month of the quarter and then declining dramatically in the first month of the next.

What these numbers could not communicate when presented as text in a table, which our brains interpret through the use of verbal processing, becomes visible and understandable when communicated visually. This is the power of "data visualization."

Although data visualization usually features relationships between quantitative values, it can also display relationships that are not quantitative in nature. For instance, the connections between people on a social networking site such as Facebook or between suspected terrorists can be displayed using a node and link visualization. In the following example, people are the nodes, represented as circles, and their relationships are the links, represented as lines that connect them.

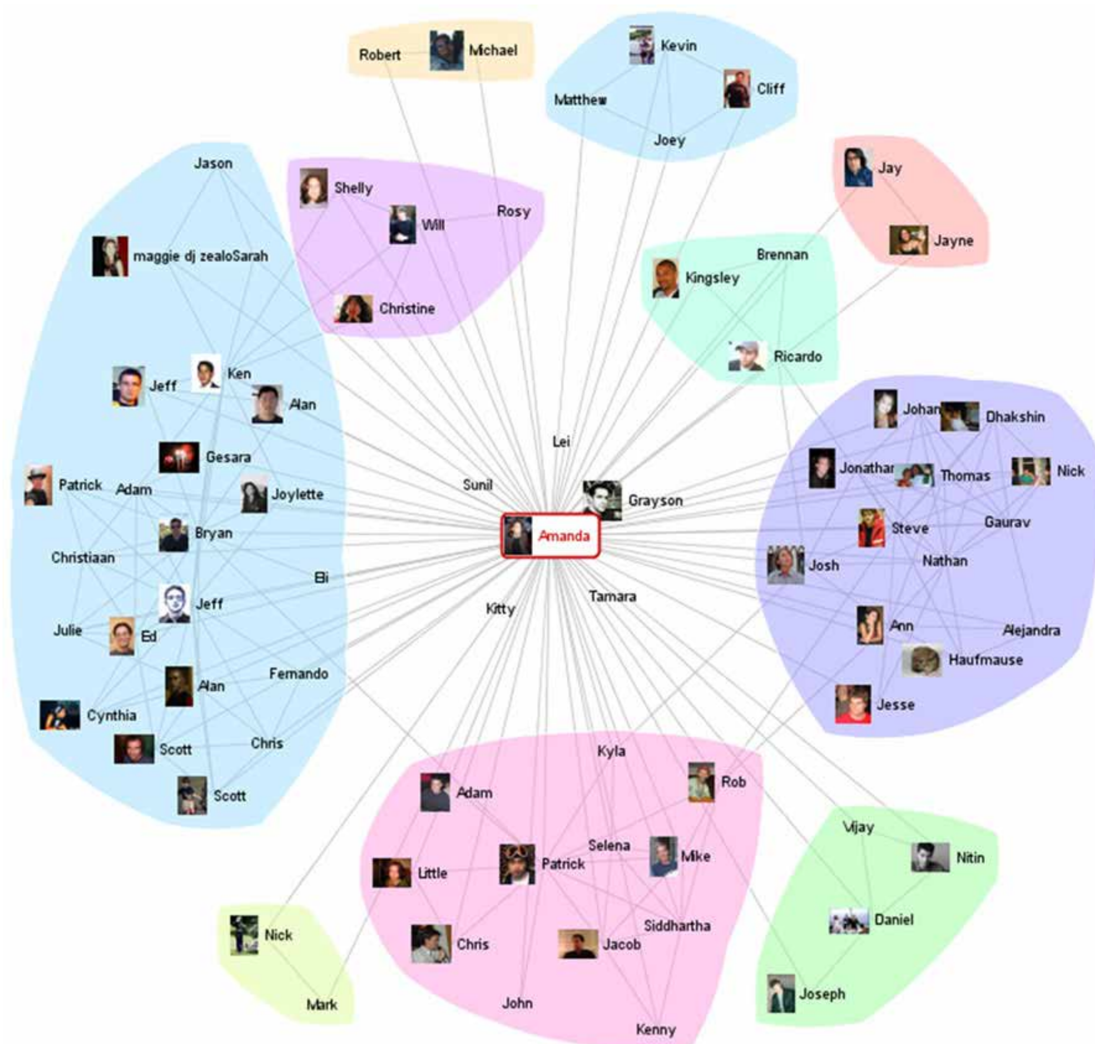



FIGURE 35.3

Visualizations that feature relationships between entities, such as the people in the example above, can be enriched with the addition of quantitative information as well. For example, the number of times that any two people have interacted could be represented by the thickness of the line that connects them.

## 35.1 DATA VISUALIZATION IN HISTORICAL CONTEXT

People have been arranging data into tables (columns and rows) at least since the 2nd century C.E., but the idea of representing quantitative information graphically didn't arise until the 17th century. For this innovation we have the French philosopher and mathematician Rene Descartes to thank. He developed a two-dimensional coordinate system for displaying values, consisting of a horizontal axis for one variable and a vertical axis for another, primarily as a graphical means of performing mathematical operations. It wasn't until the late 18th century that we began to exploit the potential of graphics for the communication of quantitative data, for which we have the Scotsman William Playfair to thank. Playfair pioneered many of the graphs that are commonly used today. He was the first person to use a line moving up and down as it progressed from left to right to show how values changed through time, as in the example below. He also invented the bar graph, and on one of his off days he invented the pie chart, which we have since found relatively ineffective, because it encodes values as visual attributes (primarily the area of each slice as well as the angle that it forms in the center of the pie) that we cannot easily perceive and compare.

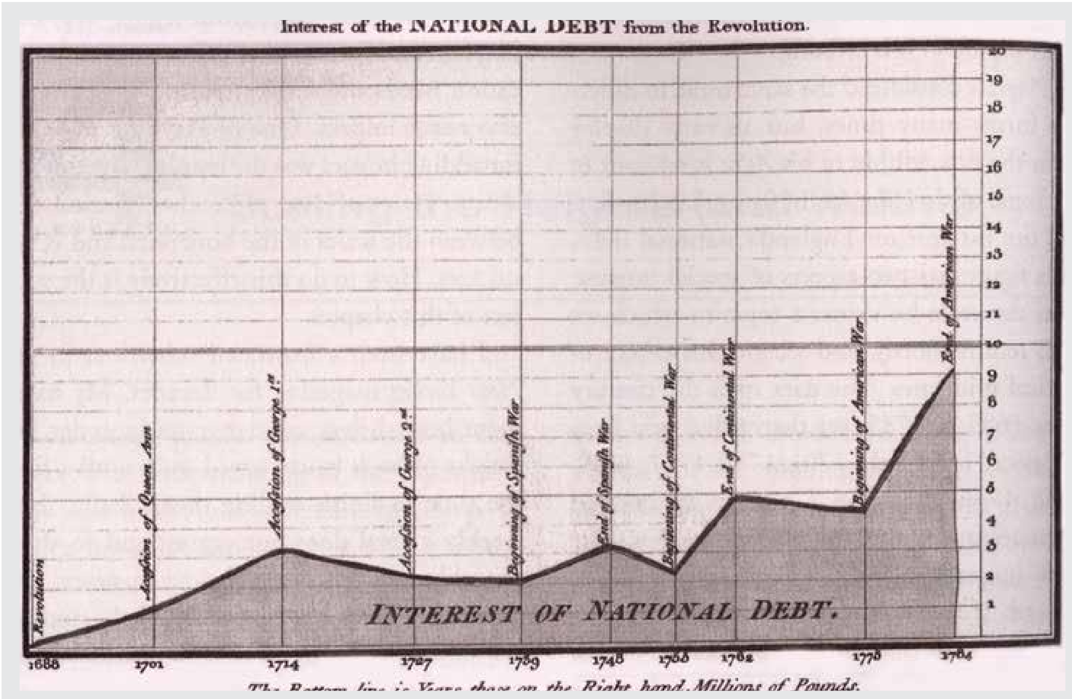



Figure 35.4: Playfair included this graph in his *The Commercial and Political Atlas* (1786) to argue against England's policy of financing colonial wars through national debt.

*Courtesy of William Playfair (1759-1823). Copyright: pd (Public Domain (information that is common property and contains no original authorship)).*

The use of quantitative graphs gradually increased over the years, but their methods and effectiveness evolved little until the second half of the 20th century. Jacques Bertin laid the foundation for much of the progress that's been made during the last half a century with the publication in 1967 of the book *Semiologie graphique* (*The Semiology of Graphics*, Bertin 1967). His work was pivotal because he discovered that visual perception operated according to rules that could be followed to express information visually in ways that represented it intuitively, clearly, accurately, and efficiently.

The person who really introduced us to the power of data visualization as a means for exploring and making sense of quantitative data was the Princeton statistics professor John Tukey, who in 1977 gave form to a whole new statistical approach called *exploratory data analysis*.

In 1983, the person working in the field today whose name is recognized above all others, Edward Tufte, published his groundbreaking book *The Visual Display of Quantitative Information*. In it he pointed out that there were effective ways of displaying data visually and then there were the ways that most people were doing it, which didn't work very well. Also working to improve data visualization practices around this time was William Cleveland, who extended and refined data visualization techniques for statisticians.

Soon thereafter, a new research specialty emerged in the academic world, which was coined "information visualization." In their 1999 book *Readings in Information Visualization: Using Vision to Think*, Stuart Card, Jock Mackinlay, and Ben Shneiderman collected the best academic work that had been done by that time into a single volume and made its discoveries accessible beyond the walls of academia (Card et al 1999).

Since the turn of the 21st century, data visualization has been popularized, too often in tragically ineffective ways as it has reached the masses through commercial software products. Gratefully, amongst the bevy of products that promote data visualization in ways that feature superficially appealing aesthetics above useful and effective data exploration, sense-making, and communication, there are a few serious contenders for our attention who are helping us fulfill its potential in practical and powerful ways.

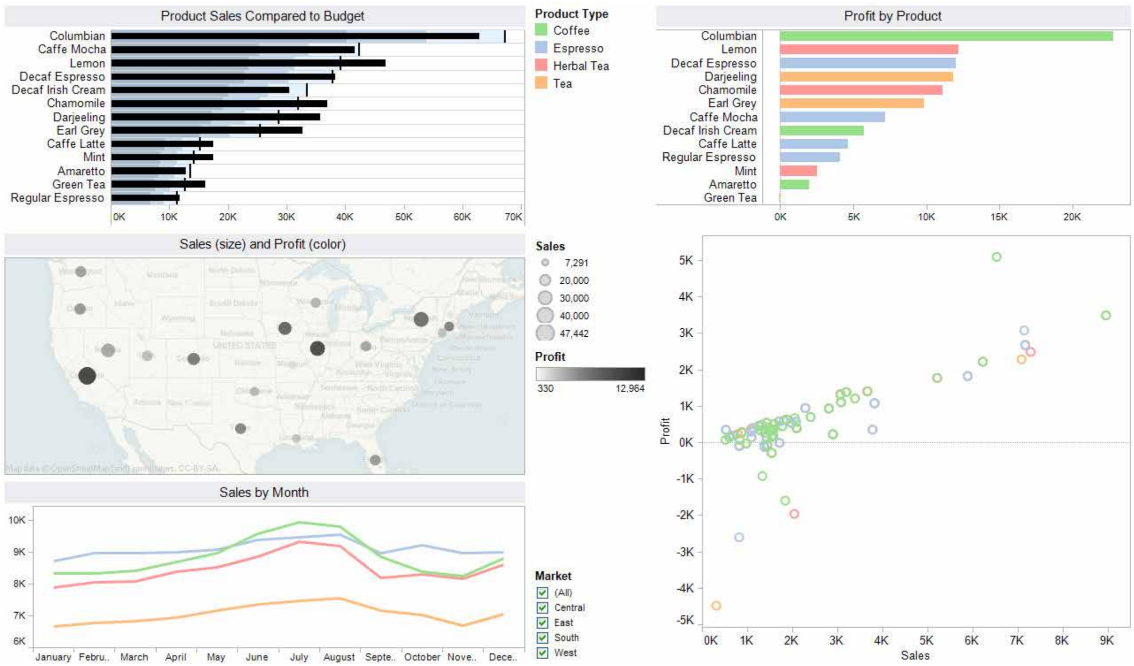



FIGURE 35.5: This display, consisting of multiple views of the same data set, was created using Tableau Software, one of the few software vendors that currently understand data visualization.



Among those who have contributed to our understanding of data visualization, Colin Ware has done the most to base its practice on an understanding of human perception. Ware's two excellent books - *Information Visualization: Perception for Design* (Ware, 2004) and *Visual Thinking for Design* (Ware 2008) - compile, organize, and explain what we have learned from several scientific disciplines about visual thinking and cognition and apply that knowledge to data visualization.

## 35.2 PICTURES FOR THE EYES AND MIND

Data visualization is only successful to the degree that it encodes information in a manner that our eyes can discern and our brains can understand. Getting this right is much more a science than an art, which we can only achieve by studying human perception. The goal is to translate abstract information into visual representations that can be easily, efficiently, accurately, and meaningfully decoded. Consider a case when you need to help people understand the primary causes of death in America contained in the following table:

| Causes of Death | Deaths per Year |
|---|---|
| Heart disease | 616,067 |
| Cancer | 562,875 |
| Stroke (cerebrovascular diseases) | 135,952 |
| Chronic lower respiratory diseases | 127,924 |
| Accidents (unintentional injuries) | 123,706 |
| Alzheimer's disease | 74,632 |
| Diabetes | 71,382 |
| Influenza and Pneumonia | 52,717 |
| Nephritis, nephrotic syndrome, and nephrosis | 46,448 |
| Septicemia | 34,828 |
| All other causes | 577,181 |
| Total | 2,423,712 |

FIGURE 35.6



To achieve this goal, the display should achieve the following:

- Clearly indicates how the values relate to one another, which in this case is a part-to-whole relationship - the number of deaths per cause, when summed, equal all deaths during the year.
- Represents the quantities accurately.
- Makes it easy to compare the quantities.
- Makes it easy to see the ranked order of values, such as from the leading cause of death to the least.

- Makes obvious how people should use the information - what they should use it to accomplish - and encourages them to do this.

The traditional way to display this information graphically involves a pie chart, illustrated below.

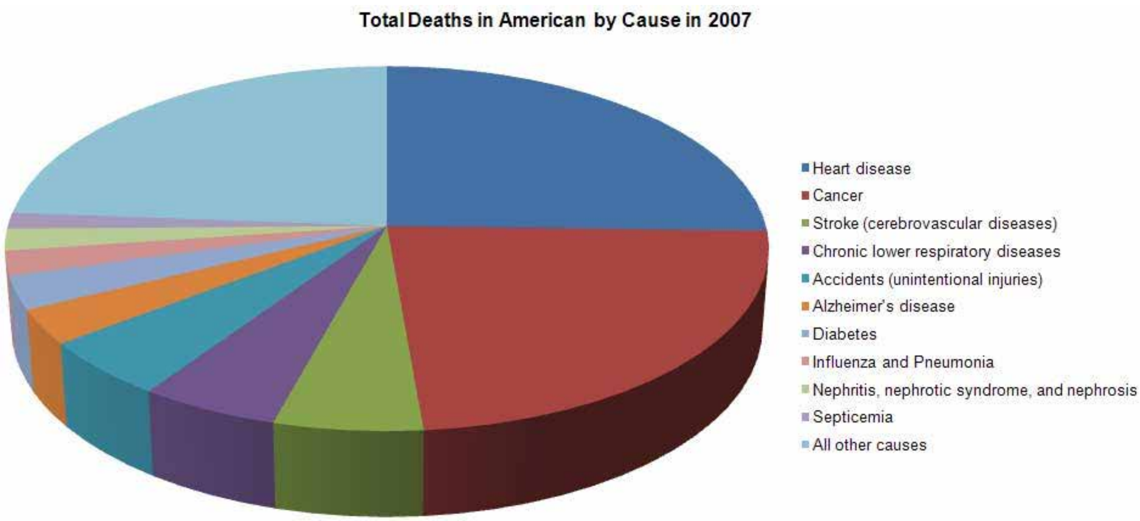



Figure 35.7

*Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.*

How well does this pie chart satisfy our criteria for effectiveness? Let's consider each of the requirements.

*Clearly indicates the nature of the relationship? Yes.* The primary strength of a pie chart is the fact that it clearly indicates a part-to-whole relationship between the values.

*Represents the quantities accurately? No.* Pie charts encode values redundantly through the use of three visual attributes: the area of each slice, the angle formed by each slice at the center of the pie, and the length of the each slice along the pie's perimeter. Even when the area, angle, and perimeter of each slice is calculated properly, it fails in that we cannot perceive any one of these attributes accurately. Visual perception in humans has not evolved to support accurate decoding of areas, angles, or distance along a curve.

*Makes it easy to compare the quantities? No*. Because we cannot perceive the values accurately, we also cannot compare them easily or accurately. Furthermore, in this particular pie chart, because a legend has been used to label the slices, we are forced over and over to look up the meaning of the slices we wish to compare by finding the right color, which is often difficult to discriminate. The fact that this pie chart has been rendered in 3-D also complicates the simple act of comparison because the perspective skews the relative size and shape of the slices, making slices on the bottom appear larger and more salient than similarly sized slices on the top.

*Makes it easy to see the ranked order of values? No*. Even though the slices are displayed in ranked order from the highest value (heart disease) at the top and continuing clockwise to the smallest, excluding the final "All other causes" slice, this ranking isn't obvious, because it's difficult to compare the slices. For example, the red cancer slice appears to be larger than the blue heart disease slice due to the 3-D effect, which has given it more visual weight. Effects such as the 3-D rendering of this pie chart are sometimes used to intentionally mislead.

*Makes obvious how people should use the information? Partially*. Although the pie chart succeeds in encouraging people to compare the slices to understand the relative contributions of each part to the whole, it fails to support this operation effectively.

Given the ways in which this pie chart has failed to match human perception, let's consider an alternative form of display. The following bar graph displays the same set of values, but in a way that can be more readily perceived.

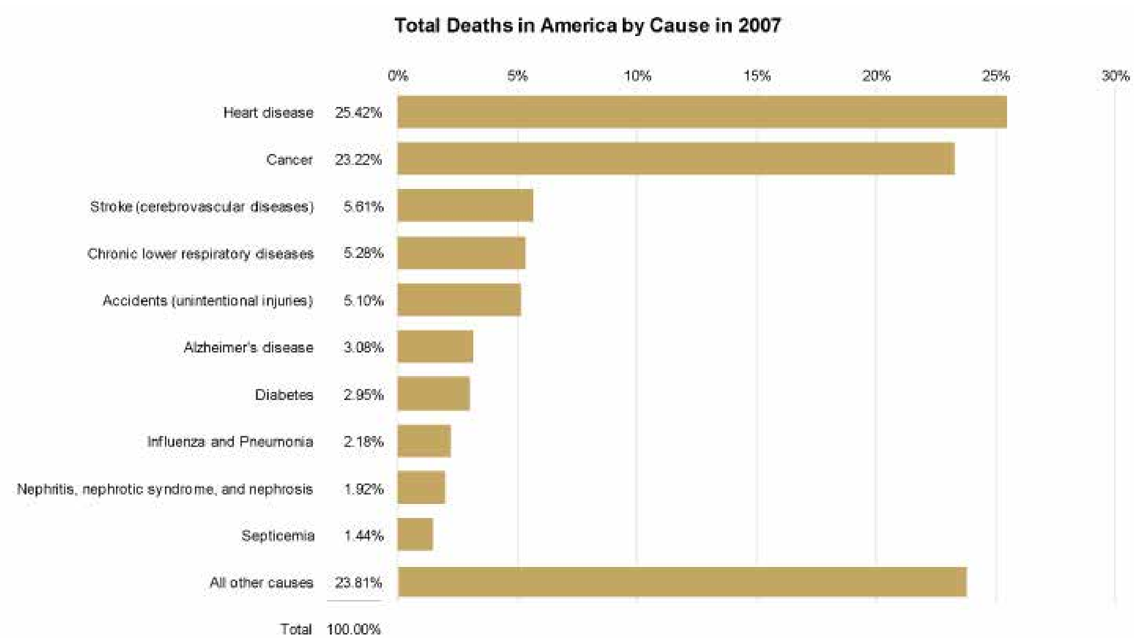



Figure 35.8



Let's review the effectiveness of this bar graph using the same criteria as before.

*Clearly indicates the nature of the relationship? Yes.* In and of itself, a bar graph does not declare the part-to-whole nature of the relationship between these values, because, unlike pie charts, bar graphs can be used to display other relationships as well. This particular bar graph, however, includes components that make the nature of the relationship clear, including the title ("Total Deaths…") and especially the column of values that add up to 100%.

*Represents the quantities accurately? Yes.* The horizontal position at which each bar ends and the length in relation to the quantitative scale along the x-axis both encode these values in a way that can be accurately perceived. Unlike areas, angles, and the lengths of curved lines that don't share a common baseline, 2-D position and the length of straight linear objects such as these bars, which share a common baseline and run parallel to one another are visual attributes that we can perceive with a high degree of accuracy.

*Makes it easy to compare the quantities? Yes*. Because we can perceive these values accurately when encoded as bars, it is also quite easy to compare them. Notice how easy it is to see differences in the lengths of these bars that could not be easily seen when comparing the slices of the pie. Also notice that when each bar shares the same color, unlike the pie's slices, which varied in color, our eyes are encouraged to compare the bars because of that likeness. And because the bars are labeled directly with the names of the causes of death, we must no longer do the work that a legend requires when comparing the values.

*Makes it easy to see the ranked order of values? Yes*. Because differences in the bar's lengths are easy to perceive, the fact that they are ranked from highest to lowest, except for the final "All other causes" bar, is obvious. By arranging the bars in ranked order, we've also made comparisons much easier by placing those causes of death that are closest in value near one another in the graph.

*Makes obvious how people should use the information? Yes*. The fact that these bars should be compared to understand the varying degree to which these causes of death contribute to total deaths is intuitively obvious.

The point of comparing the perceptual effectiveness of the pie chart and bar graph has not been to make a case against pie charts (although this case deserves to be made), but to illustrate how **we should always judge a visualization's merits by the degree to which we can easily, efficiently, accurately, and meaningfully perceive the story that the information has to tell**. To do this, we must understand the perceptual strengths and weakness of various graphical means for displaying particular stories. To do this, we must understand perception.

## 35.3 DATA VISUALIZATION AND HUMAN PERCEPTION

Data visualization is effective because it shifts the balance between perception and cognition to take fuller advantage of the brain's abilities. Seeing (i.e visual perception) which is handled by the visual cortex located in the rear of the brain, is extremely fast and efficient. We see immediately, with little effort. Thinking (i.e. cognition), which is handled primarily by the cerebral cortex in the front of the brain, is much slower and less efficient. Traditional data sensemaking and presentation methods require conscious thinking for almost all of the work. Data visualization shifts the balance toward greater use of visual perception, taking advantage of our powerful eyes whenever possible.

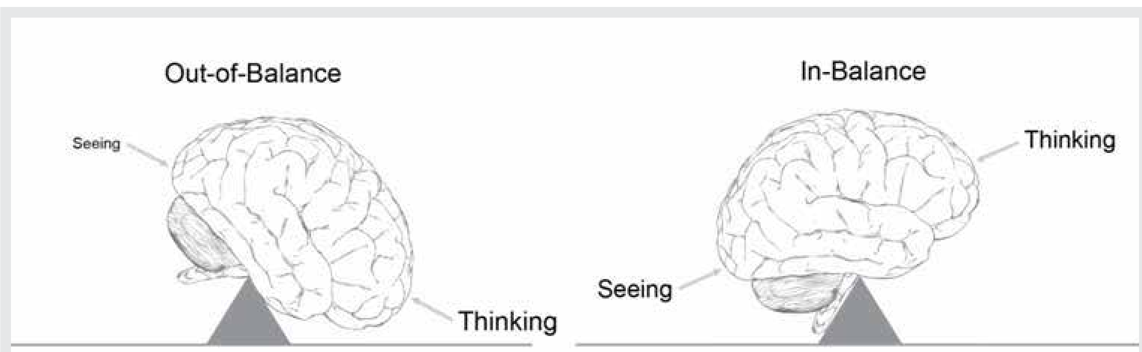
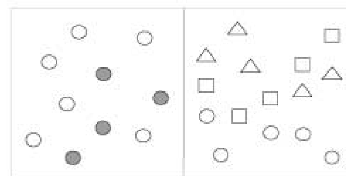
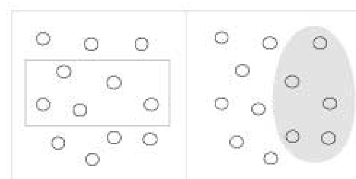
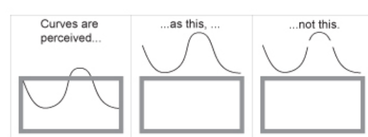
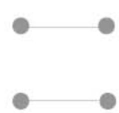



Figure 35.9



One of the earliest contributions to the science of perception was made by the Gestalt School of Psychology. The original intent of this effort when it began in 1912 was to uncover how we perceive pattern, form, and organization in what we see. The founders observed that we organize what we see in particular ways in an effort to make sense of it. The result of the effort was a series of Gestalt principles of perception, which are still respected today as accurate descriptions of visual behavior. Here are a few of the principles that can inform our data visualization efforts:

| | | |
|---|---|---|
| Proximity | Objects that are close together are perceived as a group. |  |
| Similarity | Objects that share similar attributes (e.g., color or shape) are perceived as a group. |  |
| Enclosure | Objects that appear to have a boundary around them (e.g., formed by a line or area of common color) are perceived as a group. |  |
| Closure | Open structures are perceived as closed, complete, and regular whenever there is a way that they can be reasonably interpreted as such. |  |
| Continuity | Objects that are aligned together or appear to be a continuation of one another are perceived as a group. |  |
| Connection | Objects that are connected (e.g., by a line) are perceived as a group. |  |

New insights into visual perception and cognition are arising from work in various disciplines besides information visualization, such as human factors and human-computer interaction, but none are more ground-breaking than those arising from the cognitive sciences, especially cognitive psychology. Today, with new and improved technologies and methodologies for brain exploration, opportunities to improve the perceptual effectiveness of data visualization abound. Two areas of study in particular are especially useful:

- preattentive visual processing
- mechanisms and limitations of attention and memory

One of the great strengths of data visualization is our ability to process visual information much more rapidly than verbal information. Preattentive visual processing is that part that automatically occurs in the brain prior to conscious awareness. It consists of several stages, each handled by specialized neurons that are tuned to detect particular attributes of the visual information contained in light that reflects off the surfaces of objects in the world, which is then stitched together into a picture in our mind's eye of that object. We can use these basic attributes, such as differences in length, size, hue, color intensity, angle, texture, shape, and so on, as the building blocks of data visualization. When we do so in an informed manner, we have the ability to transfer much of the work that is needed to decode the contents of a visual display, such as a graph, from the slower conscious, energy intensive parts of the brain to the faster parts of the brain that require less energy, which results in more efficient cognition.

Studies in attention and memory are revealing our surprisingly limited ability to hold multiple items simultaneously in awareness. This recognition leads us to augment attention and memory by relying on external forms of information storage. One of the most powerful ways to do this is to encode information visually, which allows more information to be chunked together into the limited slots available in working memory. Another method is to place several views of information in front of our eyes at one time, thus extending our ability to explore data multidimensional and from multiple perspectives to make comparisons and see connections to a degree that would be impossible if we had to consume these views one at a time, due to the limits of working memory. Good data visualization techniques and technologies, properly used, can extend our thinking into new realms of analytical sensemaking, and we are still only beginning to tap into this potential.

## 35.4 FUTURE DIRECTIONS

What's most needed in the field of data visualization, as in other fields, is not always what's most exciting or not even what's particularly innovative. Sometimes we simply need to make it easier to do those things that work. One example of this is the effort of a few software vendors to build data visualization best practices right into the tools, such as in the form of defaults, thereby making it easier and less time-consuming to do what works and harder and more costly to do what doesn't. Besides these simple, straightforward but often overlooked improvements, a few other areas offer the potential for enrichment, such as the following:

- The integration of geo-spatial and network displays (such as node and link diagrams) with other forms of display for seamless interaction and simultaneous use.
- Technological support for collaborative data sensemaking to bring the complementary advantage of multiple brains together.
- The application of data visualization beyond descriptive statistics to the realm of predictive analytics, such as through the use of interactive predictive visual models,
- Tighter integration of data mining algorithms to find meaningful

patterns with data visualization to provide a better way to review and explore those patterns.

- Improved human-computer interface devices for interacting with data visualization in a more rapid and seamless manner.

All of these are being pursued to some degree, but could be exploited more quickly if more researchers focused on solving real problems that we face in the world today.

## 35.5 WHERE TO LEARN MORE

Several universities have developed graduate programs that are dedicated to the study and advancement of data visualization. The University of Maryland, Stanford, the University of North Carolina, the University of California, Berkeley, and Georgia Tech are a few of the finest. Although several periodicals in the broader fields of computer graphics and human-computer interaction include articles about data visualization, only one academic journal features the field exclusively: *Information Visualization Journal*, published quarterly by Palgrave Macmillan. A few smaller publications focus on making data visualization practical and accessible to a broader audience, such as the *Visual Business Intelligence Newsletter*. Conferences dedicated to the field are also few. The oldest, IEEE's *VisWeek*, which includes the InfoVis and VAST (Visual Analytics Science and Technology) sub-conferences that are dedicated entirely to data visualization, remains the largest and perhaps best of the conferences, but significant work in the field also appears in other conferences of broader perspective, such as CHI (Computer-Human Interaction) and SIGGRAPH.

### CHI - Human Factors in Computing Systems

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 2011 | 2010 | 2009 | 2008 | 2007 | 2006 | 2005 | 2004 |
| 2003 | 2002 | 2001 | 2000 | 1999 | 1998 | 1997 | 1996 |

1995 1994 1993 1992 1991 1990 1989 1988
1987 1986 1985 1983 1982

### SIGGRAPH - International Conference on Computer Graphics and Interactive Techniques

2002 2001 2000 1999 1998 1997 1996 1995
1994 1993 1992 1991 1990 1989 1988 1987
1986 1985 1984 1983 1982 1981 1980 1979
1978 1977 1976 1975 1974

Next conference is coming up 21 Jul 2013 in Anaheim, California

### InfoVis - IEEE Symposium on Information Visualization

2005 2004 2003 2002 2001 2000 1999 1998
1997 1995

Refreshing exceptions, including Tableau Software and TIBCO Spotfire, both spin-offs of academic work, SAS JMP, which arose from a deep understanding of statistics, and a few other relatively small vendors, are gradually stealing the attention they deserve from the big software companies - especially business intelligence vendors - that dominate the market. Apart from product vendors, a few research laboratories and consultancies are also contributing to the development and application of the field, including Microsoft Research, Pacific Northwest National Laboratory, Flowing Media, Oculus Info, and Perceptual Edge.

Several good books have been written about data visualization. The following, in chronological order, are especially useful for surveying the field and as a source of basic instruction:

| | |
|---|---|
| Tufte, Edward R. (1983): *The Visual Display of Quantitative Information*. Cheshire, CT, Graphics Press | All four of Tufte's books are exceptional, but his first is the best. It makes an inspiring case for graphical excellence. |
| Cleveland, William S. (1994): *The Elements of Graphing Data*. Hobart Press | Data visualization practices focused on the needs of statisticians. |
| Harris, Robert L. (2000): *Information Graphics: A Comprehensive Illustrated Reference*. Oxford University Press, USA | An encyclopedic reference for information graphics. |
| Card, Stuart K., Mackinlay, Jock D. and Shneiderman, Ben (eds.) (1999): *Readings in Information Visualization: Using Vision to Think*. Academic Press | An overview of the best academic research in the field as of the publication date. |
| Few, Stephen (2004): *Show Me the Numbers: Designing Tables and Graphs to Enlighten*. Analytics Press | An accessible, practical, and comprehensive guide to the design of tables and graphs for communication. |
| Ware, Colin (2008): *Visual Thinking: for Design*. Morgan Kaufmann | An eloquent introduction to visual perception and cognition as it relates to data visualization. |
| Few, Stephen (2009): *Now You See It: Simple Visualization Techniques for Quantitative Analysis*. Analytics Press | An accessible and practical guide to data visualization for analysis. |

Many blogs and online discussion forums feature data visualization - some thoughtfully, based on expertise, and some with the shallowness that is often found on the Web. Here are a few of the best:

- Tufte.com (Edward Tufte)
- Perceptual Edge (Stephen Few)
- Eager Eyes (Robert Kosara)
- Visual Complexity (Manuel Lima)
- Flowing Data (Nathan Yau)
- Pictures of Numbers (Mike Dickison)
- Instant Cognition (Clint Ivy)

## 35.6 COMMENTARY BY RONALD A. RENSINK

How to cite this commentary in your report

### Ronald A. Rensink

I am interested in vision—the various ways that humans, animals, and computers use light to see. I believe that vision involves constraints that apply to any system, and that the most successful visual systems are based on very general information-processing strategies. As such, my approach is to examine biological systems (including humans) to see how they operate, and then to look a...

Ronald A. Rensink
Ronald A. Rensink is a member of The Interaction Design Foundation

### 35.6.1 Four Futures and a History

Stephen Few provides a nice overview of the reasons why we should design data visualizations to be effective, and why it's important to understand human perception when doing so. In fact, he's done this so well that I can't add much to his arguments. But I can, however, push the basic message a bit further, out into the times before and after those he discusses. Out into areas that are not as well known, or not really developed, where new opportunities and new dangers may lie...

Perhaps the best place to begin is the beginning. Discussing the beginning of visualization is not without its problems, if only for the fact that there exist several different kinds of visualization - for example, data visualization, information visualization, and scientific visualization. But whatever adjective used, we generally find a history more extensive than commonly imagined. For example, although Descartes did contribute to the graphic display of quantitative data in the 17th century, graphs had already been used to represent things such as temperature and light intensity three centuries earlier. Indeed, as Manfredo Massironi discusses in his book (Massironi, 2002; p. 131), quantities such as displacement were graphed as a function of time as far back as the 11th century. But while these facts may be of interest in their own right, the more important point is that techniques in graphic representation have been developed over many centuries, and many of these techniques have been subsequently forgotten - perhaps fallen out of vogue, or never found wide use to begin with. But the reasons for their dismissal may not necessary apply in this day and age. Indeed, several techniques might lend themselves quite well to modern technology, and so might be worth resurrecting in one form or other. Books such as Massironi's are helpful in discovering such possibilities.

On to the future. Or more precisely, on to ways of further developing useful connections between visualization and psychology. To begin with, there is potential for considerably more integration between vision and visualization than currently exists; much more processing could be offloaded to the viewer's visual cortex. As Stephen Few mentions, one way of doing so is by making use of simple preattentive properties such as length, orientation, and hue. But recent work in vision science has shown that the preattentive level of vision contains far more visual intelligence than that. Among other things, preattentive processes can determine shadows, extract three-dimensional orientation, and link scattered elements of the image into unified groups. These abilities could be exploited in higher-powered visualizations. Another area of recent progress is our understanding of visual attention and scene perception. Our visual perception of the world seems to be based on a just-in-time architecture in which attention is directed to the right object at the right time. If the co-ordination mechanisms involved can be handled correctly, it would open up the prospect of "seeing" abstract datasets in a way that is as natural and effortless as seeing the physical world. (A brief overview of these developments and their implications can be found in Rensink, 2002.)

A related opportunity is the greater use of visual analogy (or metaphor). Here, the emphasis is no longer on bypassing conscious thought, but on using modes of thought best suited for reasoning about visuospatial objects and processes. For example, when reasoning about physical force, a highly useful metaphor is the directed line, or arrow. A more modern example is the desktop, which allows a user to reason about possible actions on their computer. As in the case of visual perception, many - if not most - developments to date have been based on a relatively shallow understanding of the mechanisms involved. But given that cognitive scientists have learned much more about metaphor, it may be time to consider its use in a more sophisticated fashion. Ultimately, visualizations might be able to create mental images that correspond in a natural way to the structure of any process or task. (For an interesting discussion of this, see Paley, 2009.)

A third direction of potential importance is the creation of more powerful evaluation methods based on the methodologies developed in experimental psychology. Psychologists have spent centuries learning what to do (and not to do) to obtain precise measurements of various aspects of human behaviour. It would be good to learn from this. Of course, some of these techniques have already been adapted to evaluation. But as in the case of cognitive and perceptual mechanisms, the transfer of knowledge here is far from complete, and there is much that could still be done. For example, consider evaluating how well a given scatterplot design conveys the correlation in a dataset. In the past, this was done by presenting the viewer with the scatterplot and asking for a numerical estimate of the (perceived) correlation. But a more powerful approach is to borrow the experimental methodology of measuring just noticeable differences (jnds): the viewer is presented with two side-by-side scatterplots, and asked to choose the more correlated one. Results based on this approach show both precision and accuracy to be specified over all correlations by two functions governed by only two parameters. As a consequence, a given scatterplot design can be completely evaluated based on just two simple measurements. (For details, see Rensink and Baldridge, 2010.)

A final direction to consider - perhaps the most challenging of all - is to develop a systematic way of ensuring that visualization designs make optimal (or at least, good) use of human perception and cognition. In theory, this could result in a “science of design”. In practice, this might not be possible, if only because the number of possible designs is so immense and our understanding of human cognition so incomplete. But it may be possible to follow the example of several other areas of design, and aim for a set of principles that would at least constrain the space of possibilities to consider. For example, constraints based on physical forces or material properties can be applied to any architectural design, determining whether or not it is viable. There is no *a priori* reason why a similar approach would not also work for visualization. The efforts of Bertin are perhaps a start in this direction, providing suggestions about the kinds of graphic representation that might be applied to various kinds of problems. Work by Tufte, Mackinlay, Ware, and others have extended this further. But however useful these suggestions are, we are still a long way from a solid foundation for thinking about effective visualizations. Many foundational issues are still poorly understood. What is really going on in a visualization? Is there a way to describe this process precisely and objectively? Is it even possible in principle to determine if a given visualization draws upon the perceptual and cognitive resources of the viewer in an optimal way? The answers to these questions and others like them will be difficult to find. But they will determine the extent to which we can enable humans and machines to best combine their respective strengths.

### 35.6.2 References

Massironi, M. (2002). *The Psychology of Graphic Images: Seeing Drawing, Communicating*. Matwah NJ: Erlbaum.

Paley, W.B. (2009). Interface and mind. *it – Information Technology*, **51**: 131-141.

Rensink, R.A. (2002). Internal vs. external information in visual perception. *Proceedings of the Second International Symposium on Smart Graphics*: 63-70. [Smart Graphics 2; Hawthorne, NY, USA.]

Rensink R.A., and Baldridge G. (2010). The perception of correlation in scatterplots. *Computer Graphics Forum*, 29: 1203-1210.

## 35.7 COMMENTARY BY NAOMI B. ROBBINS

How to cite this commentary in your report

### Naomi B. Robbins

*© Naomi B. Robbins*

Naomi B. Robbins is the author of Creating More Effective Graphs, published by John Wiley (2005). She is a consultant, keynote speaker, and seminar leader who specializes in the graphical display of data. She trains employees of corporations and organizations on the effective presentation of data. She also reviews documents and presentations for clients, suggesting improvements or altern...

Naomi B. Robbins
Naomi B. Robbins is a member of The Interaction Design Foundation

Stephen Few wrote an excellent description of data visualization and the necessity for designing graphics to take advantage of our knowledge of human perception and cognition. In this commentary I question who is responsible for the myriad of visualizations that ignore this knowledge: the software vendors, the software users or others? In addition, I point out important work that deserves greater exposure on the integration of geo-spatial and other forms of data display, a topic on Few's most-needed list. I end with additional sources for learning more.

### 35.7.1 Responsibility for perceptual problems with many data visualizations

Few's article states:

"Since the turn of the 21st century, data visualization has been popularized, too often in tragically ineffective ways as it has reached the masses through commercial software products."

Certainly, software vendors are responsible for offering many graph forms that hinder rather than help the reader to understand the data. The vendors offer graphs to wow the audience rather than to communicate clearly and they create demand for ineffective graphs. But they are not solely responsible for the myriads of graphs with perceptual problems.

People learn from what they see and they see many ineffective graphs. The software users then demand software that allows them to imitate these ineffective designs. This gets us in a chicken and egg situation: Do vendors produce these awful visualizations because their customers demand them, or do the customers become attracted to them when they see what vendors market?

An example of the ineffective ways includes pseudo-third dimensions in bar charts. Figure 1 shows a pseudo-three-dimensional bar chart in Excel. Almost no one reads it correctly. I describe other problems with this graph in *Creating More Effective Graphs* [1].

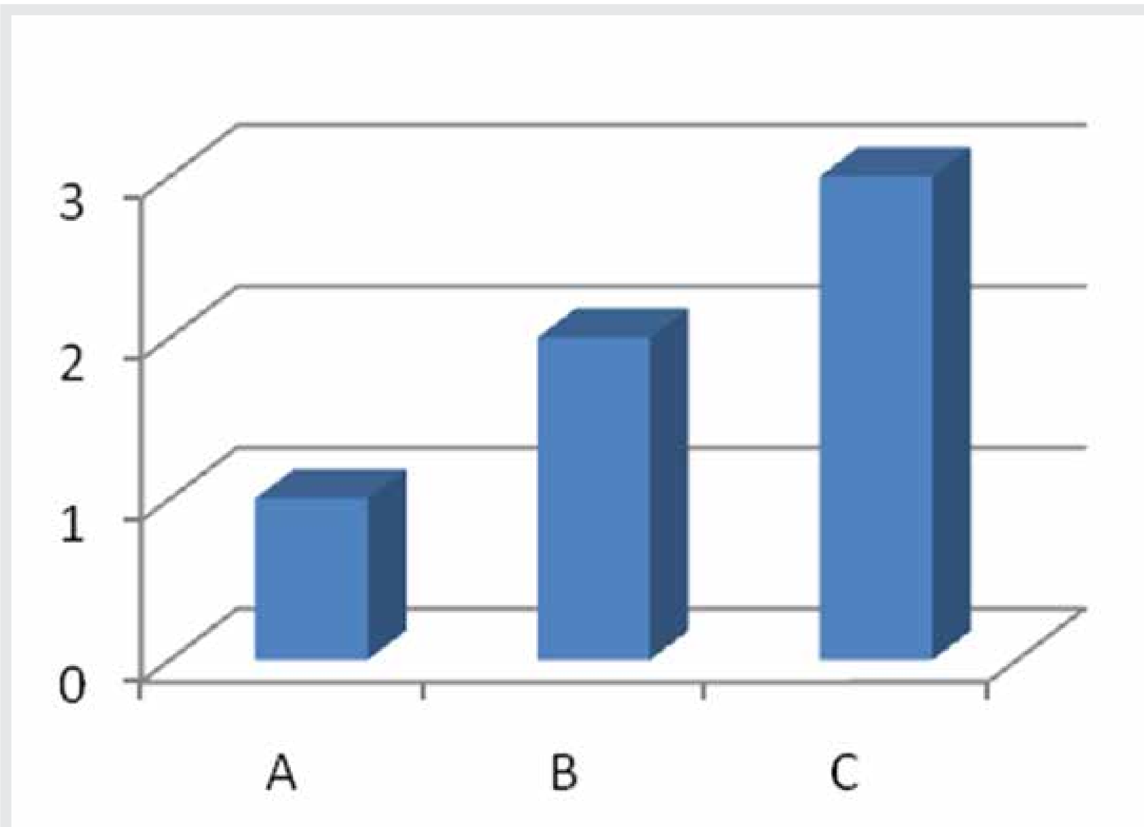



Figure 35.1: Almost no one reads this simple chart correctly. The numbers plotted are 1, 2, and 3. Plot it yourself in Excel if you don't believe me.



A number of graphic artists have made major contributions to the field of data visualization. However, there are some graphic artists who have no appreciation of numbers and don't realize that the representation of numbers in graphs should be proportional to the numbers they represent. As a result, it is common to see graphs that are not drawn to scale.

Some graph designers want to give the impression of better performance than is actually the case and intentionally design graphs that mislead to achieve this impression. Other graph designers may be more concerned with demonstrating their technological abilities or artistic abilities than in communicating clearly and accurately. Until recently, our educational system did not provide training in communicating numbers. Today, there are some excellent courses at the college level but the majority of people receive little, if any, training in presenting numerical information. Therefore, many graph designers are unaware of the principles of effective graphs. Some of the problems occur from a lack of proofreading and careless errors.

As an analogy, a current style in fashion is high-heeled shoes. A quick search on "dangers of high heels" revealed that there has been an increase in the number of bunion operations on wearers of high heels as well as foot pain, back pain and neck pain. In some cases the Achilles tendon grows shorter. Balance is affected so that the risk of falls is greater. The list of problems goes on and on. Is the shoe designer, the shoe manufacturer, the retail outlet that sells the shoes or the customer who buys them responsible for this increase in medical problems? Is this situation analogous to the data visualization one? Both cause serious problems: poor business decisions in one case and pain and suffering as well as unnecessary medical expenses in the other. I hope that these questions stimulate interesting discussion.

### 35.7.2 Integration of geo-spatial displays with other forms of display

In his section on future directions, Few mentions areas that offer the potential for enrichment including the integration of geo-spatial displays with other forms of display for seamless interaction and simultaneous use. Several researchers have made advances in this area. For example, the micromap designs of Dan Carr [1] and [2] add a geographic context to statistical information, allowing for the joint exploration of statistical and geographic patterns in data. As illustrated in Figure 2, statistical graphics, here dots, are linked to small maps by color. In the first row, we can see that Maryland is represented by red dots and so Maryland is shaded red on the right-hand map. Sorting by poverty level, we see that not only are poverty and education inversely related, but that there is a geographic clustering of southern U.S. states by these variables.

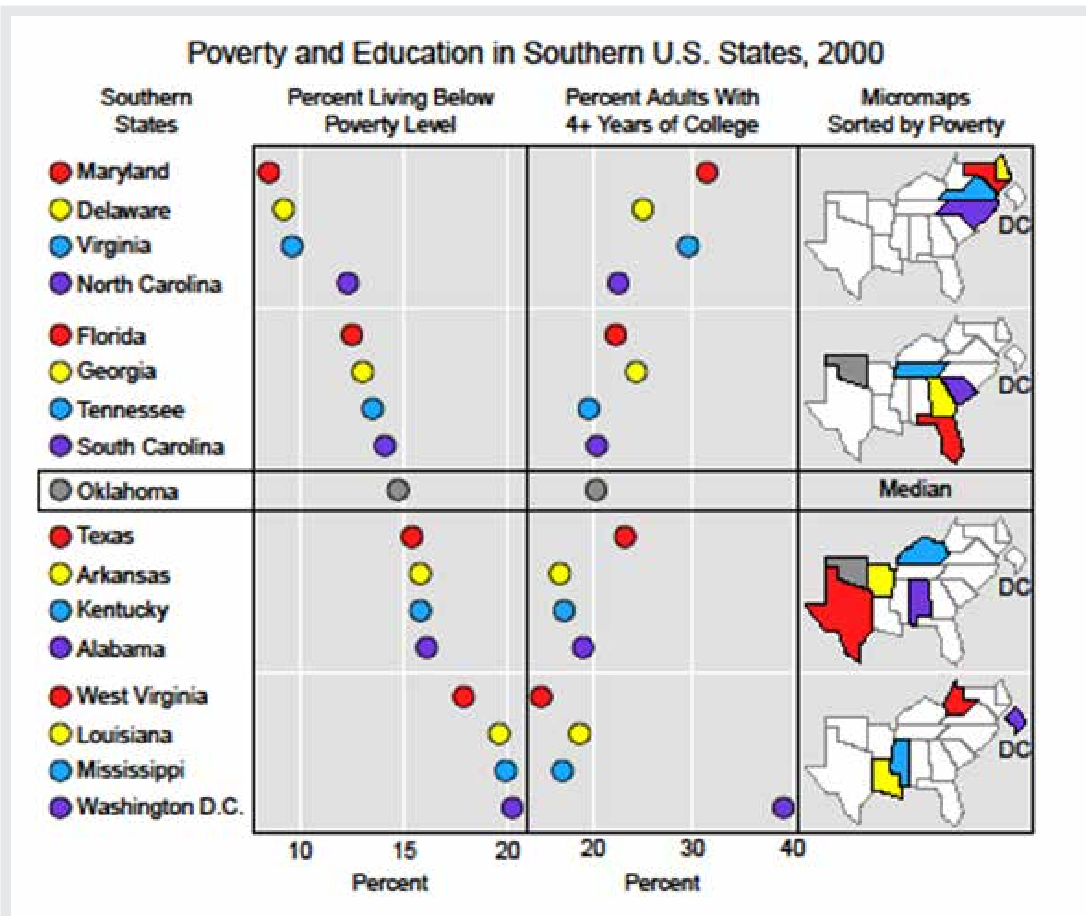



Figure 35.2: An example of a micromap design from Carr and Pickle [1].

### 35.7.3 Where to Learn More

Data visualization does not belong to a single academic discipline. Statisticians, computer scientists, psychologists, graphic designers and others practice and contribute to data visualization. The university programs and resources that Few mentions lean heavily towards computer science. A few excellent programs joining statistical graphics with computer science are available at George Mason University, Iowa State, and the University of Augsburg. There are many others. I will leave it to other commentators to add excellent programs in cognitive psychology and graphic design. The Journal of Computational and Graphical Statistics, a joint publication of the American Statistical Association, the Institute of Mathematical Statistics and the Interface Foundation of North America is another academic journal on the topic. The Statistical Computing Statistical Graphics Newsletter (SCGN) is another informal publication. Although the Joint Statistical Meetings are not exclusively devoted to statistical graphics and data visualization, there are as many sessions sponsored by the Statistical Graphics Section as many a smaller conference contains.

One addition I would make to the “what’s needed” list is better communication between the computer scientists, graphic designers, psychologists and statisticians. More joint conferences and attending each other’s conferences would help each discipline benefit from the research of the others.

### 35.7.4 References

Carr, Daniel B. and Linda Williams Pickle. 2010. *Visualizing Data Patterns with Micromaps*. Chapman and Hall/CRC, Boca Raton, FL.

Carr DB, Wallin JF and Carr DA. Two new templates for epidemiology applications: Linked micromap plots and conditioned choropleth maps. *Statistics in Medicine* 19:2521-38, 2000.

Robbins, Naomi B. 2005. *Creating More Effective Graphs*. John Wiley and Sons, Hoboken, NJ.

## 35.8 COMMENTARY BY ROBERT KOSARA

How to cite this commentary in your report

### Robert Kosara

© ***Robert Kosara***

I am an Assistant Professor at the Department of Computer Science, College of Information Technology, at the University of North Carolina at Charlotte (UNCC), where I am also a member of the Charlotte Visualization Center. I received both my Ph.D. (2001) and M.S. degrees from Vienna University of Technology (Vienna, Austria). Before coming to Charlotte, I worked at the VRVis Research ...

Robert Kosara
Robert Kosara is a member of The Interaction Design Foundation

### 35.8.1 Metaphors and Interaction

One important topic Stephen Few only mentions briefly in his very well-written and comprehensive piece is interaction. While static charts and visualizations are undoubtedly useful, they make little use of the immense computing power that is readily available to us today. Interaction in visualization enables the fast exploration and discovery of data patterns that the user may not even have expected. It is also possible to reduce the amount of data shown at the same time, providing clearer visualizations, while still giving the user the option to get that information on demand at any time.

Ben Shneiderman captured the role of interaction in his famous visual information seeking mantra (Shneiderman, 1996): *overview first, zoom and filter, then details on demand*. Abstract information spaces require an overview so the user has an idea where to even find data, but then it is necessary to zoom in to see details. Filtering data is important when dealing with larger datasets. Finally, details on what is shown (and also what is not shown) can be retrieved by the user as needed. All of these steps require interaction, where the user tells the visualization what he or she wants to see.

### 35.8.2 Simple Interactions

Among the simplest interactions are tooltips or other data displays that appear when the user points at a part of a visualization. Take the causes of death bar chart in Few's article above: the numbers could be shown purely on demand, perhaps including not just percentage but total number. Also, a vertical line could be drawn from the end of the active bar to the scale at the top, to make it easier to see the bars in context.

This type of interaction is effortless and easy to discover: just move your mouse over the display and see if anything happens. Displaying numbers in charts is also rather common. But the real power comes from the more advanced interactions.

### 35.8.3 Linking and Brushing

*Brushing* lets the user selects data points that get highlighted in one or more views of the same data. When several views are involved, the fact that all of them highlight the same data points is commonly referred to as *linking* (and the views are called *coordinated multiple views*). Consider this example of linked bar charts of data about passengers on the Titanic. Each bar chart represents one data dimension (class, gender, age, and survived), and shows a histogram of how many people were in each of the categories.


FIGURE 35.1

*Courtesy of Robert Kosara. Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.*

To find out how many people survived in each category, we will select the relevant bar, which will brush those data points in all the views. We can now compare survival rates for different sexes, classes, etc. by looking at how much of their respective bars is highlighted.

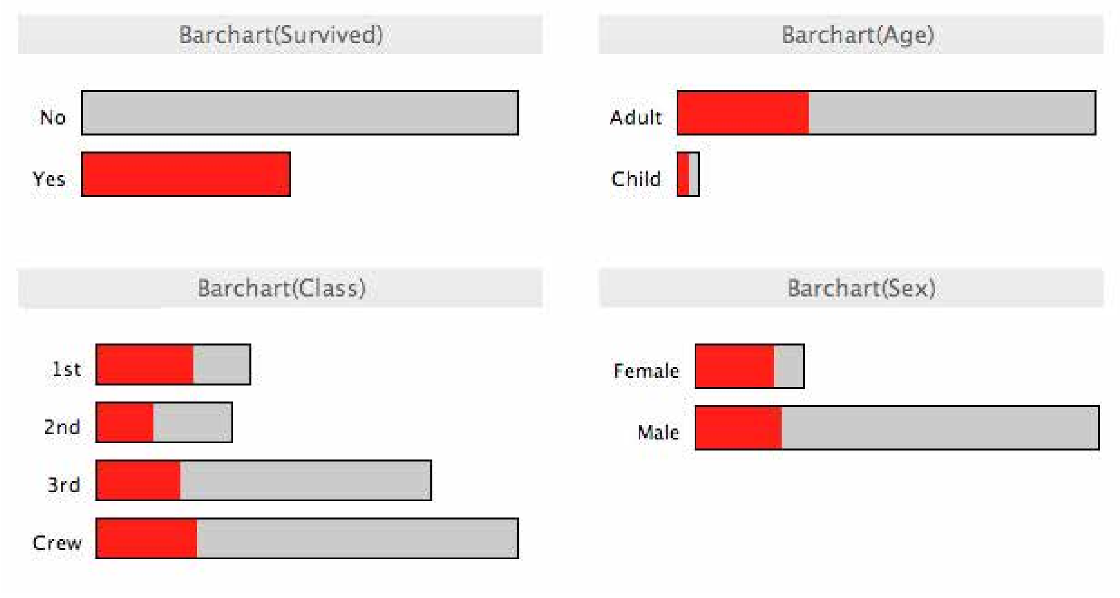



FIGURE 35.2

*Courtesy of Robert Kosara. Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms.*

The mechanism is very similar for individual data points rather than summary data like in this example. Brushing and linking make it possible to find out high-dimensional relationships in the data by trying out different possibilities.

### 35.8.4 Metaphors and Structure

Metaphors have a somewhat complicated history in visualization. There is not even a clear understanding what a metaphor even is: many people talk about visual metaphors when they mean different ways of depicting data, but others use them specifically for somewhat embellished visualizations (flowers growing to represent traffic in chat rooms, etc.).

What I want to add here is a combination of both, perhaps best summarized as structure: how do the relationships between elements in the visualization influence how people read the data? Caroline Ziemkiewicz and I have done work on this topic, and have found that the big-picture structure plays a bigger role than most people would assume.

When comparing different types of tree visualizations, we found that different studies had come to different conclusions as to which method works better based on which metaphor was used in the question: A being contained in B, or A being below B in the hierarchy. We did a study and found that there was, indeed, a compatibility effect between the linguistic metaphor used in the question and the visual metaphor of the visualization (Ziemkiewicz and Kosara, 2008).

We recently showed that there is an apparent effect of gravity between objects in a visualization that can distort the perception of distance (Ziemkiewicz and Kosara, 2010).

### 35.8.5 The Future

While we know a lot about how to create reasonable visualizations, there is still a lot we do not know or are not yet aware of. Even seemingly basic knowledge like how the layout of a visualization influences our reading of the data still needs more work to be understood and turned into useful recommendations and best practices.

Interaction is not exactly a new topic in visualization research, but is still rather rudimentary in many visualization and charting programs. To really unlock the power of visualization, these programs will need more advanced capabilities as well as ways to educate their users about their interactive features. Visualization has a lot more to offer than what most people are aware of today.

### 35.8.6 References

- Shneiderman, Ben (1996): The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. In: VL 1996 1996. pp. 336-343.

- Caroline Ziemkiewicz, Robert Kosara, The Shaping of Information by Visual Metaphors, Transactions on Visualization and Computer Graphics (Proceedings InfoVis), vol. 14, no. 6, pp. 1269-1276, 2008.
- Caroline Ziemkiewicz, Robert Kosara, Laws of Attraction: From Perceived Forces to Conceptual Similarity, Transactions on Visualization and Computer Graphics (Proceedings InfoVis), vol. 16, no. 6, pp. 1009-1016, 2010.

## 35.9 REFERENCES

*Card, Stuart K., Mackinlay, Jock D. and Shneiderman, Ben (eds.) (1999): Readings in Information Visualization: Using Vision to Think. Academic Press*

*Cleveland, William S. (1994): The Elements of Graphing Data. Hobart Press*

*Few, Stephen (2009): Now You See It: Simple Visualization Techniques for Quantitative Analysis. Analytics Press*

*Few, Stephen (2004): Show Me the Numbers: Designing Tables and Graphs to Enlighten. Analytics Press*

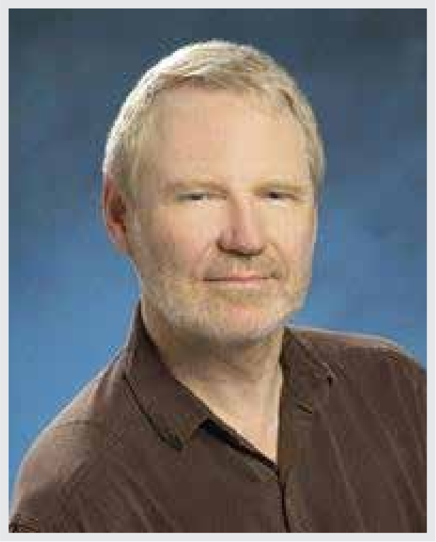
*Harris, Robert L. (2000): Information Graphics: A Comprehensive Illustrated Reference. Oxford University Press, USA*

*Tufte, Edward R. (1983): The Visual Display of Quantitative Information. Cheshire, CT, Graphics Press*

*Ware, Colin (2008): Visual Thinking: for Design. Morgan Kaufmann*

# About the author

## Stephen Few



© *Stephen Few*

Has also published under the name of:
"Steve Few"

Personal Homepage:
http://www.perceptualedge.com/

Current place of employment:
Perceptual Edge

Stephen Few has over 20 years of experience as an innovator, consultant, and educator in the fields of business intelligence (a.k.a. data warehousing and decision support) and information design. Through his company, Perceptual Edge, he focuses on the effective analysis and presentation quantitative business information. Stephen is recognized as a world leader in the field of data visualization. He teaches regularly at conferences such as those presented by The Data Warehousing Institute (TDWI) and DCI, and also in the MBA program at the Haas School of Business at U. C. Berkeley. He is also the author of the book "Show Me the Numbers: Designing Tables and Graphs to Enlighten" (Analytics Press).

## Your notes and thoughts on chapter 35

Record your notes and thoughts on this chapter. If you want to share these thoughts with others online, go to the bottom of the page at:
http://www.interaction-design.org/encyclopedia/data_visualization_for_human_perception.html

**NOTES:**

# The Creative Commons Attribution -NoDerivs 3.0 Unported License

THE WORK (AS DEFINED BELOW) IS PROVIDED UNDER THE TERMS OF THIS CREATIVE COMMONS PUBLIC LICENSE ("CCPL" OR "LICENSE"). THE WORK IS PROTECTED BY COPYRIGHT AND/OR OTHER APPLICABLE LAW. ANY USE OF THE WORK OTHER THAN AS AUTHORIZED UNDER THIS LICENSE OR COPYRIGHT LAW IS PROHIBITED.

BY EXERCISING ANY RIGHTS TO THE WORK PROVIDED HERE, YOU ACCEPT AND AGREE TO BE BOUND BY THE TERMS OF THIS LICENSE. TO THE EXTENT THIS LICENSE MAY BE CONSIDERED TO BE A CONTRACT, THE LICENSOR GRANTS YOU THE RIGHTS CONTAINED HERE IN CONSIDERATION OF YOUR ACCEPTANCE OF SUCH TERMS AND CONDITIONS.

**1. Definitions**

a. **"Adaptation"** means a work based upon the Work, or upon the Work and other pre-existing works, such as a translation, adaptation, derivative work, arrangement of music or other alterations of a literary or artistic work, or phonogram or performance and includes cinematographic adaptations or any other form in which the Work may be recast, transformed, or adapted including in any form recognizably derived from the original, except that a work that constitutes a Collection will not be considered an Adaptation for the purpose of this License. For the avoidance of doubt, where the Work is a musical work, performance or phonogram, the synchronization of the Work in timed-relation with a moving image ("synching") will be considered an Adaptation for the purpose of this License.

b. **"Collection"** means a collection of literary or artistic works, such as encyclopedias and anthologies, or performances, phonograms or broadcasts, or other works or subject matter other than works listed in Section 1(f) below, which, by reason of the selection and arrangement of their contents, constitute intellectual creations, in which the Work is included in its entirety in unmodified form along with one or more other contributions, each constituting separate and independent works in themselves, which together are assembled into a collective whole. A work that constitutes a Collection will not be considered an Adaptation (as defined above) for the purposes of this License.

c. **"Distribute"** means to make available to the public the original and copies of the Work through sale or other transfer of ownership.

d. **"Licensor"** means the individual, individuals, entity or entities that offer(s) the Work under the terms of this License.

e. **"Original Author"** means, in the case of a literary or artistic work, the individual, individuals, entity or entities who created the Work or if no individual or entity can be identified, the publisher; and in addition (i) in the case of a performance the actors, singers, musicians, dancers, and other persons who act, sing, deliver, declaim, play in, interpret or otherwise perform literary or artistic works or expressions of folklore; (ii) in the case of a phonogram the producer being the person or legal entity who first fixes the sounds of a performance or other sounds; and, (iii) in the case of broadcasts, the organization that transmits the broadcast.

f. **"Work"** means the literary and/or artistic work offered under the terms of this License including without limitation any production in the literary, scientific and artistic domain, whatever may be the mode or form of its expression including digital form, such as a book, pamphlet and other writing; a lecture, address, sermon or other work of the same nature; a dramatic or dramatico-musical work; a choreographic work or entertainment in dumb show; a musical composition with or without words; a cinematographic work to which are assimilated works expressed by a process analogous to cinematography; a work of drawing, painting, architecture, sculpture, engraving or lithography; a photographic work to which are assimilated works expressed by a process analogous to photography; a work of applied art; an illustration, map, plan, sketch or three-dimensional work relative to geography, topography, architecture or science; a performance; a broadcast; a phonogram; a compilation of data to the extent it is protected as a copyrightable work; or a work performed by a variety or circus performer to the extent it is not otherwise considered a literary or artistic work.

g. **"You"** means an individual or entity exercising rights under this License who has not previously violated the terms of this License with respect to the Work, or who has received express permission from the Licensor to exercise rights under this License despite a previous violation.

h. **"Publicly Perform"** means to perform public recitations of the Work and to communicate to the public those public recitations, by any means or process, including by wire or wireless means or public digital performances; to make available to the public Works in such a way that members of the public may access these Works from a place and at a place individually chosen by them; to perform the Work to the public by any means or process and the communication to the public of the performances of the Work, including by public digital performance; to broadcast and rebroadcast the Work by any means including signs, sounds or images.

i. **"Reproduce"** means to make copies of the Work by any means including without limitation by sound or visual recordings and the right of fixation and reproducing fixations of the Work, including storage of a protected performance or phonogram in digital form or other electronic medium.

**2. Fair Dealing Rights.** Nothing in this License is intended to reduce, limit, or restrict any uses free from copyright or rights arising from limitations or exceptions that are provided for in connection with the copyright protection under copyright law or other applicable laws.

**3. License Grant.** Subject to the terms and conditions of this License, Licensor hereby grants You a worldwide, royalty-free, non-exclusive, perpetual (for the duration of the applicable copyright) license to exercise the rights in the Work as stated below:

a. to Reproduce the Work, to incorporate the Work into one or more Collections, and to Reproduce the Work as incorporated in the Collections; and,

b. to Distribute and Publicly Perform the Work including as incorporated in Collections.

c. For the avoidance of doubt:

i. **Non-waivable Compulsory License Schemes**. In those jurisdictions in which the right to collect royalties through any statutory or compulsory licensing scheme cannot be waived, the Licensor reserves the exclusive right to collect such royalties for any exercise by You of the rights granted under this License;

ii. **Waivable Compulsory License Schemes**. In those jurisdictions in which the right to collect royalties through any statutory or compulsory licensing scheme can be waived, the Licensor waives the exclusive right to collect such royalties for any exercise by You of the rights granted under this License; and,

iii. **Voluntary License Schemes**. The Licensor waives the right to collect royalties, whether individually or, in the event that the Licensor is a member of a collecting society that administers voluntary licensing schemes, via that society, from any exercise by You of the rights granted under this License.

The above rights may be exercised in all media and formats whether now known or hereafter devised. The above rights include the right to make such modifications as are technically necessary to exercise the rights in other media and formats, but otherwise you have no rights to make Adaptations. Subject to Section 8(f), all rights not expressly granted by Licensor are hereby reserved.

**4. Restrictions.** The license granted in Section 3 above is expressly made subject to and limited by the following restrictions:

a. You may Distribute or Publicly Perform the Work only under the terms of this License. You must include a copy of, or the Uniform Resource Identifier (URI) for, this License with every copy of the Work You Distribute or Publicly Perform. You may not offer or impose any terms on the Work that restrict the terms of this License or the ability of the recipient of the Work to exercise the rights granted to that recipient under the terms of the License. You may not sublicense the Work. You must keep intact all notices that refer to this License and to the disclaimer of warranties with every copy of the Work You Distribute or Publicly Perform. When You Distribute or Publicly Perform the Work, You may not impose any effective technological measures on the Work that restrict the ability of a recipient of the Work from You to exercise the rights granted to that recipient under the terms of the License. This Section 4(a) applies to the Work as incorporated in a Collection, but this does not require the Collection apart from the Work itself to be made subject to the terms of this License. If You create a Collection, upon notice from any Licensor You must, to the extent practicable, remove from the Collection any credit as required by Section 4(b), as requested.

b. If You Distribute, or Publicly Perform the Work or Collections, You must, unless a request has been made pursuant to Section 4(a), keep intact all copyright notices for the Work and provide, reasonable to the medium or means You are utilizing: (i) the name of the Original Author (or pseudonym, if applicable) if supplied, and/or if the Original Author and/or Licensor designate another party or parties (e.g., a sponsor institute, publishing entity, journal) for attribution ("Attribution Parties") in Licensor's copyright notice, terms of service or by other reasonable means, the name of such party or parties; (ii) the title of the Work if supplied; (iii) to the extent reasonably practicable, the URI, if any, that Licensor specifies to be associated with the Work, unless such URI does not refer to the copyright notice or licensing information for the Work. The credit required by this Section 4(b) may be implemented in any reasonable manner; provided, however, that in the case of a Collection, at a minimum such credit will appear, if a credit for all contributing authors of the Collection appears, then as part of these credits and in a manner at least as prominent as the credits for the other contributing authors. For the avoidance of doubt, You may only use the credit required by this Section for the purpose of attribution in the manner set out above and, by exercising Your rights under this License, You may not implicitly or explicitly assert or imply any connection with, sponsorship or endorsement by the Original Author, Licensor and/or Attribution Parties, as appropriate, of You or Your use of the Work, without the separate, express prior written permission of the Original Author, Licensor and/or Attribution Parties.

c. Except as otherwise agreed in writing by the Licensor or as may be otherwise permitted by applicable law, if You Reproduce, Distribute or Publicly Perform the Work either by itself or as part of any Collections, You must not distort, mutilate, modify or take other derogatory action in relation to the Work which would be prejudicial to the Original Author's honor or reputation.

**5. Representations, Warranties and Disclaimer**

UNLESS OTHERWISE MUTUALLY AGREED TO BY THE PARTIES IN WRITING, LICENSOR OFFERS THE WORK AS-IS AND MAKES NO REPRESENTATIONS OR WARRANTIES OF ANY KIND CONCERNING THE WORK, EXPRESS, IMPLIED, STATUTORY OR OTHERWISE, INCLUDING, WITHOUT LIMITATION, WARRANTIES OF TITLE, MERCHANTIBILITY, FITNESS FOR A PARTICULAR PURPOSE, NONINFRINGEMENT, OR THE ABSENCE OF LATENT OR OTHER DEFECTS, ACCURACY, OR THE PRESENCE OF ABSENCE

OF ERRORS, WHETHER OR NOT DISCOVERABLE. SOME JURISDICTIONS DO NOT ALLOW THE EXCLUSION OF IMPLIED WARRANTIES, SO SUCH EXCLUSION MAY NOT APPLY TO YOU.

**6. Limitation on Liability.** EXCEPT TO THE EXTENT REQUIRED BY APPLICABLE LAW, IN NO EVENT WILL LICENSOR BE LIABLE TO YOU ON ANY LEGAL THEORY FOR ANY SPECIAL, INCIDENTAL, CONSEQUENTIAL, PUNITIVE OR EXEMPLARY DAMAGES ARISING OUT OF THIS LICENSE OR THE USE OF THE WORK, EVEN IF LICENSOR HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES.

**7. Termination**

a. This License and the rights granted hereunder will terminate automatically upon any breach by You of the terms of this License. Individuals or entities who have received Collections from You under this License, however, will not have their licenses terminated provided such individuals or entities remain in full compliance with those licenses. Sections 1, 2, 5, 6, 7, and 8 will survive any termination of this License.

b. Subject to the above terms and conditions, the license granted here is perpetual (for the duration of the applicable copyright in the Work). Notwithstanding the above, Licensor reserves the right to release the Work under different license terms or to stop distributing the Work at any time; provided, however that any such election will not serve to withdraw this License (or any other license that has been, or is required to be, granted under the terms of this License), and this License will continue in full force and effect unless terminated as stated above.

**8. Miscellaneous**

a. Each time You Distribute or Publicly Perform the Work or a Collection, the Licensor offers to the recipient a license to the Work on the same terms and conditions as the license granted to You under this License.

b. If any provision of this License is invalid or unenforceable under applicable law, it shall not affect the validity or enforceability of the remainder of the terms of this License, and without further action by the parties to this agreement, such provision shall be reformed to the minimum extent necessary to make such provision valid and enforceable.

c. No term or provision of this License shall be deemed waived and no breach consented to unless such waiver or consent shall be in writing and signed by the party to be charged with such waiver or consent.

d. This License constitutes the entire agreement between the parties with respect to the Work licensed here. There are no understandings, agreements or representations with respect to the Work not specified here. Licensor shall not be bound by any additional provisions that may appear in any communication from You. This License may not be modified without the mutual written agreement of the Licensor and You.

e. The rights granted under, and the subject matter referenced, in this License were drafted utilizing the terminology of the Berne Convention for the Protection of Literary and Artistic Works (as amended on September 28, 1979), the Rome Convention of 1961, the WIPO Copyright Treaty of 1996, the WIPO Performances and Phonograms Treaty of 1996 and the Universal Copyright Convention (as revised on July 24, 1971). These rights and subject matter take effect in the relevant jurisdiction in which the License terms are sought to be enforced according to the corresponding provisions of the implementation of those treaty provisions in the applicable national law. If the standard suite of rights granted under applicable copyright law includes additional rights not granted under this License, such additional rights are deemed to be included in the License; this License is not intended to restrict the license of any rights under applicable law.

ii. The Interaction Design Foundation Addendum to the Creative Commons licence

The Interaction Design Foundation Addendum to the Creative Commmons licence is a placeholder for additions to the Creative Commons licence, which are deemed necessary to include in consideration of Danish law and the operation of this site and The Interaction Design Foundation.

**1. Attribution**

If this work is used under the licencing conditions set forth here, attribution must be clearly given, i.e. the author's name, the title and URL of this work/publication/web page must clearly appear. The attribution must be given in a manner appropriate to the medium in which it is given: For example, electronic copies must include a clickable URL, which does not use the nofollow attribute value.

**2. Updates**

Internet technology, publishing technology, and the applicable laws, rules, and regulations change frequently. Accordingly, The Interaction Design Foundation reserves the unilateral right to update, modify, change and alter its Site Terms and Conditions as well as Copyright Terms at any time. All such updates, modifications, changes and alterations are binding on all users and browsers of Interaction-Design.org, readers of electronic and non-eletronic versions of the publications produced by The Interaction Design Foundation. Such updates will be posted on Interaction-Design.org.

iii. Exceptions

## Exceptions

Many materials published by The Interaction Design Foundation - both in print and electronically - may contain materials where the copyright is owned by a third party, e.g. another publisher. In this case, the copyright status depends on the third party, i.e. the copyright owner, and may for example be “all rights reserved -

used with permission". When this is the case, we clearly label the content. For images, we both write the specific copyright label (including attribution) underneath the caption in both electronic and print copies as well as include the copyright label (including attribution) inside the image file (i.e. the full-resolution version) in metadata types like EXIF, IPTC, and XMP. We only include and label content with the following copyright terms:

a. Pd:
Public Domain (information that is common property and contains no original authorship)
Legal Code (full licence text): http://en.wikipedia.org/wiki/Public_domain

b. CompositeWorkWithMultipleCopyrightTerms:
Work that is derived from or composed of multiple works with varying copyright terms and/or copyright holders

c. FairUse:
Copyrighted materials that meet the legal criteria for Fair Use when used by the Interaction Design FoundationThe most common cases of Fair Use are: 1) Cover art: Cover art from various items, for identification only in the context of critical commentary of that item (not for identification without critical commentary). 2) Team and corporate logos: For identification. 3) Other promotional material: Posters, programs, billboards, ads: For critical commentary. 4) Film and television screen shots: For critical commentary and discussion of the cinema and television. 5) Screenshots from software products: For critical commentary. 6) Paintings and other works of visual art: For critical commentary, including images illustrative of a particular technique or school. 7) Images with iconic status or historical importance: As subjects of commentary. 8) Images that are themselves subject of commentary.
Legal Code (full licence text): http://en.wikipedia.org/wiki/Fair_use

d. AllRightsReservedUsedWithoutPermission:
All Rights Reserved. Non-free, copyrighted materials used without permission. The materials are used without permission of the copyright holder because the materials meet the legal criteria for Fair Use and/ or because The Interaction Design Foundation has not been able to contact the copyright holder. The most common cases of Fair Use are: 1) Cover art: Cover art from various items, for identification only in the context of critical commentary of that item (not for identification without critical commentary). 2) Team and corporate logos: For identification. 3) Other promotional material: Posters, programs, billboards, ads: For critical commentary. 4) Film and television screen shots: For critical commentary and discussion of the cinema and television. 5) Screenshots from software products: For critical commentary. 6) Paintings and other works of visual art: For critical commentary, including images illustrative of a particular technique or school. 7) Images with iconic status or historical importance: As subjects of commentary. 8) Images that are themselves subject of commentary.

e. AllRightsReserved:
All Rights Reserved. Materials used with permission. Permission to use has been granted exclusively to The Interaction Design Foundation and/ or the author of the given work/chapter, in which the copyrighted material is used. This permission constitutes a non-transferable license and, as such, only applies to The Interaction Design Foundation. Therefore, no part of this material may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, recording or otherwise without prior written permission of the copyright holder.

f. CC-Att-1:
Creative Commons Attribution 1.0 Unported
Legal Code (full licence text): http://creativecommons.org/licenses/

by/1.0/

g. CC-Att-3:
Creative Commons Attribution 3.0 Unported
Legal Code (full licence text): http://creativecommons.org/licenses/by/3.0/

h. CC-Att-2:
Creative Commons Attribution 2.0 Unported
Legal Code (full licence text): http://creativecommons.org/licenses/by/2.0/

i. CC-Att:
Creative Commons Attribution 3.0 Unported
Legal Code (full licence text): http://creativecommons.org/licenses/by/3.0/

j. CC-Att-ND-3:
Creative Commons Attribution-NoDerivs 3.0 Unported
Legal Code (full licence text): http://creativecommons.org/licenses/by-nd/3.0/

k. CC-Att-ND-2:
Creative Commons Attribution-NoDerivs 2.0 Unported
Legal Code (full licence text): http://creativecommons.org/licenses/by-nd/2.0/

l. CC-Att-ND-1:
Creative Commons Attribution-NoDerivs 1.0 Unported
Legal Code (full licence text): http://creativecommons.org/licenses/by-nd/1.0/

m. CC-Att-ND:

Creative Commons Attribution-NoDerivs 3.0 Unported
Legal Code (full licence text): http://creativecommons.org/licenses/by-nd/3.0/

n. CC-Att-SA-1:
Creative Commons Attribution-ShareAlike 1.0 Unported
Legal Code (full licence text): http://creativecommons.org/licenses/by-sa/1.0/

o. CC-Att-SA-3:
Creative Commons Attribution-ShareAlike 3.0
Legal Code (full licence text): http://creativecommons.org/licenses/by-sa/3.0/

p. CC-Att-SA-2:
Creative Commons Attribution-ShareAlike 2.0 Unported
Legal Code (full licence text): http://creativecommons.org/licenses/by-sa/2.0/

q. CC-Att-SA:
Creative Commons Attribution-ShareAlike 3.0 Unported
Legal Code (full licence text): http://creativecommons.org/licenses/by-sa/3.0/

r. Unknown:
Copyright status unknown

s. Trademarks and logos:
All trademarks, logos, service marks, collective marks, design rights, personality rights or similar rights that are mentioned, used or cited by The Interaction Design Foundation and its authors are the property of their respective owners. The use of any trademark in our materials does not vest in the author or The Interaction Design Foundation

# CHAPTER 38

# Human-Robot Interaction

by Kerstin Dautenhahn.

This chapter introduces and critically reflects upon some key challenges and open issues in Human-Robot Interaction (HRI) research. The chapter emphasizes that in order to tackle these challenges, both the user-centred and the robotics-centred aspects of HRI need to be addressed. The synthetic nature of HRI is highlighted and discussed in the context of methodological issues. Different experimental paradigms in HRI are described and compared. Furthermore, I will argue that due to the artificiality of robots, we need to be careful in making assumptions about the 'naturalness' of HRI and question the widespread assumption that humanoid robots should be the ultimate goal in designing successful HRI. In addition to building robots for the purpose of providing services for and on-behalf of people, a different direction in HRI is introduced, namely to use robots as social mediators between people. Examples of HRI research illustrate these ideas.

## 38.1 BACKGROUND

Human-Robot Interaction (HRI) is a relatively young discipline that has attracted a lot of attention over the past few years due to the increasing availability of complex robots and people's exposure to such robots in their daily lives, e.g. as robotic toys or, to some extent, as household appliances (robotic vacuum cleaners or lawn movers). Also, robots are increasingly being developed for real world application areas, such as robots in rehabilitation, eldercare, or robots used in robot-assisted therapy and other assistive or educational applications.

This article is not meant to be a review article of HRI per se, please consult e.g. (Goodrich and Schultz, 2007; Dautenhahn, 2007a) for such surveys and discussions of the history and origins of this field. Instead, I would like to discuss a few key issues within the domain of HRI that often lead to misunderstandings or misinterpretations of research in this domain. The chapter will not dwell into technical details but focus on interdisciplinary aspects of this research domain in order to inspire innovative new research that goes beyond traditional boundaries of established disciplines.

Researchers may be motivated differently to join the field HRI. Some may be roboticists, working on developing advanced robotic systems with possible real-world applications, e.g. service robots that should assist people in their homes or at work, and they may join this field in order to find out how to handle situations when these robots need to interact with people, in order to increase the robots' efficiency. Others may be psychologists or ethologists and take a human-centred perspective on HRI; they may use robots as tools in order to understand fundamental issues of how humans interact socially and communicate with others and with interactive artifacts. Artificial Intelligence and Cognitive Science researchers may join this field with the motivation to understand and develop complex intelligent systems, using robots as embodied instantiations and testbeds of those.

Last but not least, a number of people are interested in studying the *interaction* of people and robots, how people perceive different types and behaviours of robots, how they perceive social cues or different robot embodiments, etc. The means to carry out this work is usually via ’user studies’. Such work has often little technical content; e.g. it may use commercially available and already fully programmed robots, or research prototypes showing few behaviours or being controlled remotely (via the Wizard-of-Oz approach whereby a human operator, unknown to the participants, controls the robot), in order to create very constrained and controlled experimental conditions. Such research strongly focuses on humans’ reactions and attitudes towards robots. Research in this area typically entails large-scale evaluations trying to find statistically significant results. Unfortunately this area of ’user studies’, which is methodologically heavily influenced by experimental psychology and human-computer interaction (HCI) research, is often narrowly equated with the field of “HRI”. “Shall we focus on the AI and technical development of the robot or shall we do HRI”? is not an uncommon remark heard in research discussions. This tendency to equate HRI with ’user studies ’ is in my view very unfortunate, and it may in the long run sideline HRI and transform this field into a niche-domain. HRI as a research domain is a synthetic science, and it should tackle the whole range of challenges from technical, cognitive/AI to psychological, social, cognitive and behavioural.

## 38.2 HRI - A SYNTHETIC, NOT A NATURAL SCIENCE

HRI is a field that has emerged during the early 1990s and has been characterized as:

> *“Human—Robot Interaction (HRI) is a field of study dedicated to understanding, designing, and evaluating robotic systems for use by or with humans”*
>
> *(Goodrich and Schultz, 2007, p. 204).*

What is Human-robot interaction (HRI) and what does it try to achieve?

> *"The HRI problem is to understand and shape the interactions between one or more humans and one or more robots"*
>
> *(Goodrich and Schultz, 2007, p. 216).*

The characterization of the fundamental HRI problem given above focuses on the issues of understanding what happens between robots and people, and how these interactions can be shaped, i.e. influenced, improved towards a certain goal etc.

The above view implicitly assumes a reference point of what is meant by "robot". The term is often traced back to the Czechoslovakian word *robota* (work), and its first usage is attributed to Karel Capek's play R.U.R.: Rossum's Universal Robots (1920). However, the term "robot" is far from clearly defined. Many technical definitions are available concerning its motor, sensory and cognitive functionalities, but little is being specified about the robot's appearance, behaviour and interaction with people. As it happens, if a non-researcher interacts with a robot that he or she has never encountered before, then what matters is how the robot looks, what it does, and how it interacts and communicates with the person. The 'user' in such a context will not care much about the cognitive architecture that has been implemented, or the programming language that has been used, or the details of the mechanical design.

Behaviours and appearances of robots have dramatically changed since the early 1990s, and they continue to change — new robots appearing on the market, other robots becoming obsolete. The design range of robot appearances is huge, ranging from mechanoid (mechanical-looking) to zoomorphic (animal-looking robots) to humanoid (human-like) machines as well as android robots at the extreme end of human-likeness. Similarly big is the design space of robot appearance, behaviour and their cognitive abilities. Most robots are unique designs, their hardware and often software may be incompatible with other robots or even previous versions of the same robot. Thus, robots are generally discrete, isolated systems, they have not evolved in the same way as natural species have evolved, they have not adapted during evolution to their environments. When biological species evolve, new generations are connected to the previous generations in non-trivial ways; in fact, one needs to know the evolutionary history of a species in order to fully appreciate its morphology, biology, behaviour and other features. Robots are designed by people, and are programmed by people. Even for robots that are learning, they have been programmed how and when to learn. Evolutionary approaches to robots' embodiment and control (Nolfi and Floreano, 2000; Harvey et al., 2005) and developmental approaches to the development of a robot's social and cognitive abilities (Lungarella et al., 2003; Asada et al., 2009; Cangelosi et al., 2010; Vernon et al., 2011; Nehaniv et al., 2013) may one day create a different situation, but at present, robots used in HRI are human-designed systems. This is very different from ethology, experimental psychology etc. which study biological systems. To give an example, in 1948 Edward C. Tolman wrote his famous article "Cognitive Maps in Rats and Men". Still today his work is among the key cited articles in research on navigation and cognitive maps in humans and other animals. Rats and people are still the same two species; they have since 1948 not transformed into completely different organisms, results gained in 1948 can still be compared with results obtained today. In contrast, the robots that were available in the early 1990s and today's robots do not share a common evolutionary history; they are just very different robotic 'species'.

Thus, what we mean by 'robot' today will be very different from what we mean by 'robot' in a hundreds of year time. The concept of robot is a *moving target*, we constantly reinvent what we consider to be ’robot’. Studying interactions with robots and gaining general insights into HRI applicable across different platforms is therefore a big challenge. Focusing only on the ’H’ in HRI, ’user studies’ , i.e. the human perspective, misses the important ’R’, the robot component, the technological and robotics characteristics of the robot. Only a deep investigation of both aspects will eventually illuminate the illusive ’I’, the interaction that emerges when we put people and interactive robots in a shared context. In my perspective, the key challenge and characterization of HRI can be phrased as follows:

“HRI is the science of studying people’s behaviour and attitudes towards robots in relationship to the physical, technological and interactive features of the robots, with the goal to develop robots that facilitate the emergence of human-robot *interactions* that are at the same time efficient (according to the original requirements of their envisaged area of use), but are also acceptable to people, and meet the social and emotional needs of their individual users as well as respecting human values”.

## 38.3 HRI - METHODOLOGICAL ISSUES

As discussed in the previous section, the concept of ’robot’ is a moving target. Thus, different from the biological sciences, research in HRI is suffering from not being able to compare results directly from studies using different types of robots. Ideally, one would like to carry out every HRI experiments with a multitude of robots and corresponding behaviours — which is practically impossible.

Let us consider a thought experiment and assume our research question is to investigate how a cylindrically shaped mobile robot should approach a seated person and how the robot’s behaviour and appearance influences people’s reactions. The robot will be programmed to carry a bottle of water, approach the person from a certain distance, stop at a certain distance in the vicinity of the person, orient its front (or head) towards the person and say "Would you like a drink?". Video cameras record people's reactions to the robot, and after the experiment they complete a questionnaire on their views and experiences of the experiment. Note, there is in fact no bi-directional interaction involved, the person is mainly passive. The scenario has been simplified this way to be able to test different conditions. We only consider three values for each category, i.e. no continuous values. Despite these gross simplifications, as indicated in table 38.1 below, we will end up with $3^7 = 2187$ combinations and possible experimental conditions to which we may expose participants to. For each condition we need a number, X, of participants, in order to satisfy statistical constraints. Each session, if kept to a very minimal scenario, will take at least 15 minutes, plus another 15 minutes for the introduction, debriefing, questionnaires/interviews, as well as signing of consent forms etc. Note, more meaningful HRI scenarios, e.g. those we conduct in our Robot House described below, typically involve scheduling one full hour for each participant per session. Since people's opinions of and behaviours towards robots is likely to change in long-term interactions, each person should be exposed to the same condition 5 times, which gives 10935 different sessions. Also, the participants need to be chosen carefully, ideally one would also consider possible age and gender differences,as well as personality characteristics and other individual differences — which means repeating the experiment with different groups of participants. Regardless of whether we expose one participant to all conditions, or we choose different participants for each condition, getting sufficient data for meaningful statistical analysis will clearly be impractical. We end up with about 328050 * X minutes required for the experiment, not considering situations where the experiment has to be interrupted due to a system's failure, rescheduling of appointments for participants etc. Clearly, running such an experiment is impractical, and not desirable, given that only minimal conditions are being addressed, so results from this experiment would necessarily be very limited and effort certainly not worthwhile.

| Features | | | |
|---|---|---|---|
| Height | 2m | 1m | 50cm |
| Speed | Fast | medium | slow |
| Voice | Human-like | Robot-like | none |
| Colour of body | Red | blue | white |
| Approach distance to person | Close | medium | far |
| Approach direction to person | Frontal approach | Side approach | Side-back approach |
| Head | Head with human-like features | Mechanical head | No head |
| ... | | | |

Table 38.1: HRI thought experiment.

Given these constraints, a typical HRI experiment simplifies to an even greater extent. The above study could limit itself to a short and tall robot and two different approach distances, resulting in 4 experimental conditions. The results would indicate how robot height influences people's preferred approach distances but only in a very limited sense, since all other features would have to be held constant, i.e. the robot's appearance (apart from height), speed, voice, colour, approach direction, head feature, etc. would be chosen once and then kept constant for the whole experiment. Thus, any results from our hypothetical experiment would not allow us to extrapolate easily to other robot designs and behaviour, or other user groups. Robots are designed artifacts, and they are a moving target; what we consider to be a typical 'robot' today will probably be very different from what people in 200 years consider to be a "robot". So will the results we have gained over the past 15 or 20 years still be applicable to tomorrow's robots?

As I have pointed out previously (Dautenhahn, 2007b) HRI is often compared to other experimental sciences, such as ethology and in particular experimental, or even clinical psychology. And indeed, quantitative methods used in these domains often provide valuable guidelines and sets of established research methods that are used to design and evaluate HRI experiments, typically focusing on quantitative, statistical methods requiring large-scale experiments, i.e. involving large sample sizes of participants, and typically one or more control conditions. Due the nature of this work the studies are typically short-term, exposing participants to a particular condition only once or a few times. Textbooks on research methods in experimental psychology can provide guidelines for newcomers to the field. However, there is an inherent danger if such approaches are taken as the gold standard for HRI research, i.e. if any HRI study is measured against it. This is very unfortunate since in fact, many methodological approaches exist that provide different, but equally valuable insights into human-robot interaction. Such qualitative methods may include in-depth, long-term case studies where individual participants are exposed to robots over an extensive period of time. The purpose of such studies is more focused on the actual *meaning* of the interaction, the experience of the participants, any behavioural changes that may occur and changes in participants' attitudes towards the robots or the interaction. Such approaches often lack control conditions but analyse in great detail interactions over a longer period of time. Other approaches, e.g. conversation-analytic methods (Dickerson et al., 2013; Rossano et al., 2013) may analyse in depth the detailed nature of the interactions and how interaction partners respond and attend to each other and coordinate their actions.

In the field of assistive technology and rehabilitation robotics, where researchers develop robotic systems for specific user groups, control conditions with different user groups are usually not required: if one develops systems to assist or rehabilitate people with motor impairments after a stroke, design aids to help visually impaired people, or develop robotic technology to help with children with autism learn about social behaviour and communication, contrasting their use of a robotic system with how healthy/neurotypical people may use the same system does not make much sense. We already know about the specific impairments of our user groups, and the purpose of such work is not to highlight again how they differ from healthy/neurotypical people. Also, often the diversity of responses within the target user group is of interest. Thus, in this domain, control groups only make sense if those systems are meant to be used for different target user groups, and so comparative studies can highlight how each of them would use and could benefit (or not) from such a system. However, most assistive, rehabilitative systems are especially designed for people with special needs, in which case control conditions with different user groups are not necessarily useful.

Note, an important part of control conditions in assistive technology is to test different systems or different versions of the same system in different experimental conditions. Such comparisons are important since they a) allow gaining data to further improve the system, and b) can highlight the added value of an assistive system compared to other conventional systems or approaches. For example, Werry and Dautenhahn (2007) showed that an interactive, mobile robot engages children with autism better than a non-robotic conventional toy.

A physician or physiotherapist may use robotic technology in order to find out about the *nature* of a particular medical condition or impairment, e.g. to find out about the nature of motor impairment after stroke, and may use an assessment robot to be tested with both healthy people and stroke patients. Similarly, a psychologist may study the nature of autism by using robotic artefacts, comparing,e.g. how children respond to social cues, speech or tactile interaction. Such artifacts would be *tools* in the research on the nature of the disorder or disability, rather than an assistive tool built to assist the patients — which means it would also have to take into consideration the patient's individual differences, likes and dislikes and preferences in the context of using the tool.

Developing complex robots for human-robot interaction requires substantial amount of resources in terms of researchers, equipment, know-how, funding and it is not uncommon that the development of such a robot may take years until it is fully functioning. Examples of this are the robot 'butler' Care-O-bot® 3 (Parlitz et al, 2008; Reiser et al., 2013, cf. Fig. 1) whose first prototype was first developed as part of the EU FP6 project COGNIRON (2004-2008), or the iCub robot (Metta et al. 2010,Fig. 2) developed from 2004-2008 as part of the 5.5-year FP6 project Robotcub. Both robots are still under development and upgraded regularly. The iCub was developed as a research platform for developmental and cognitive robotics by a large consortium, concluding several European partners developing the hardware and software of the robot. Another example is the IROMEC platform that was developed from 2006-2009 as part of the FP6 project IROMEC , Fig. 3. The robot has been developed as a social mediator for children with special needs who can learn through play. Results of the IROMEC project do not only include the robotic platform, but also a framework for developing scenarios for robot-assisted play (Robins et al., 2010), and a set of 12 detailed play scenarios that the Robot-Assisted Therapy (RAT) community can use according to specific developmental and educational objectives for each child (Robins et al., 2012). In the IROMEC project a dedicated user-centred design approach was taken (Marti and Bannon, 2009; Robins et al. 2010), however time ran out at the end of the project to do a second design cycle in order to modify the platform based on trials with the targeted end-users. Such modifications would have been highly desirable, since interactions between users and new technology typically illuminate issues that have not been considered initially. In the case of the iCub the robot was developed initially as a new cognitive systems research robotics platform, so no concrete end users were envisaged. In the case of the Care-O-bot® three professional designers were involved in order to derive a 'friendly' design (Parlitz et al., 2008).



FIGURE 38.1 A-B: The Care-O-bot® 3 robot in the UH Robot House, investigating robot assistance for elderly users as part of the ACCOMPANY project (2011, ongoing). See a video (http://www.youtube.com/watch?v=qp47BPw__9M). The Robot House is based off-campus in a residential area, and is a more naturalistic environment for the study of home assistance robots than laboratory settings, cf. Figure 6. Bringing HRI into natural environments poses many challenges but also opportunities (e.g. Sabanovic et al. 2006; Kanda et al., 2007; Huttenrauch et al. 2009; Kidd and Breazeal, 2008; Kanda et al. 2010; Dautenhahn, 2007; Woods et al., 2007; Walters et al., 2008).

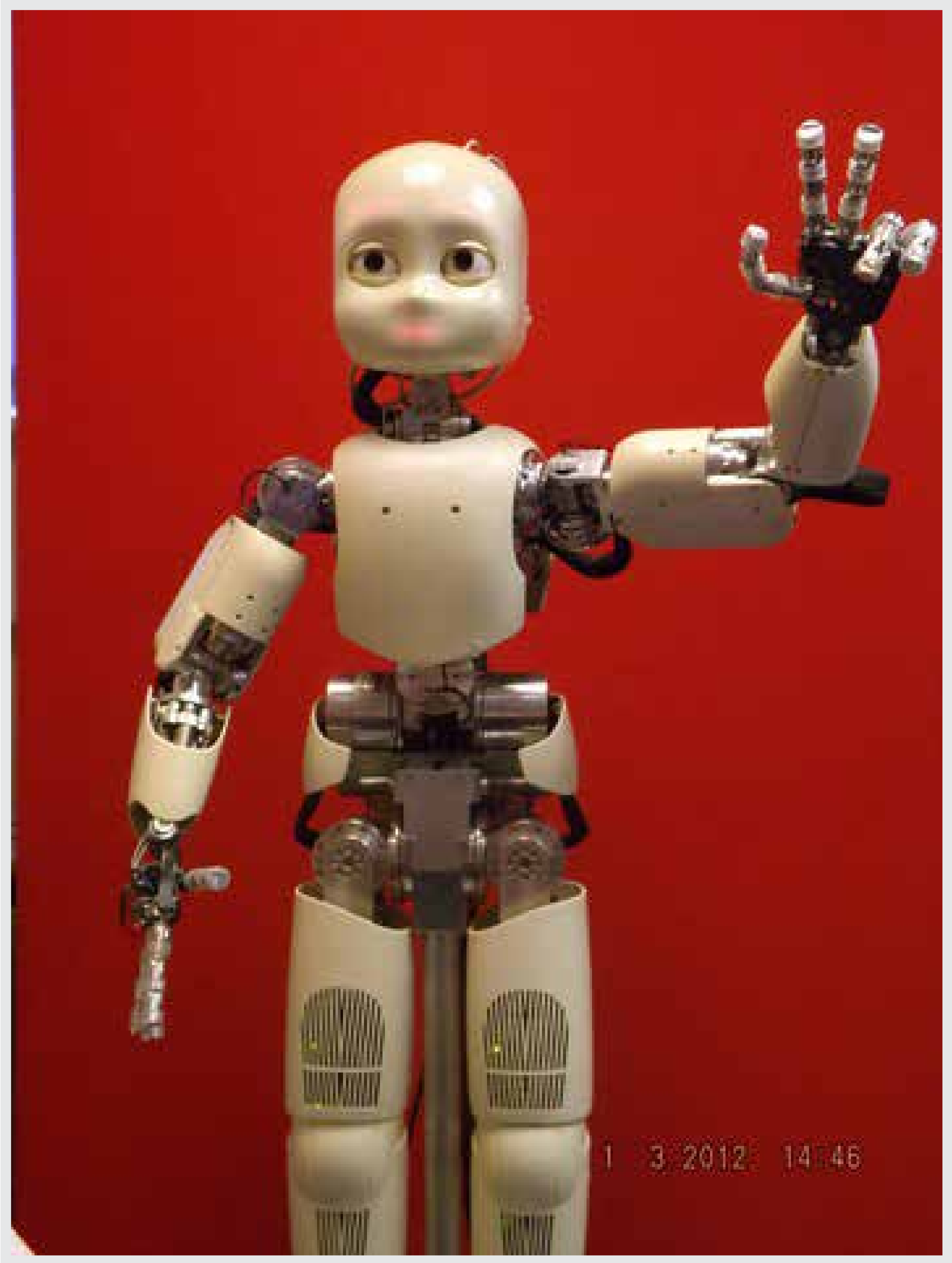


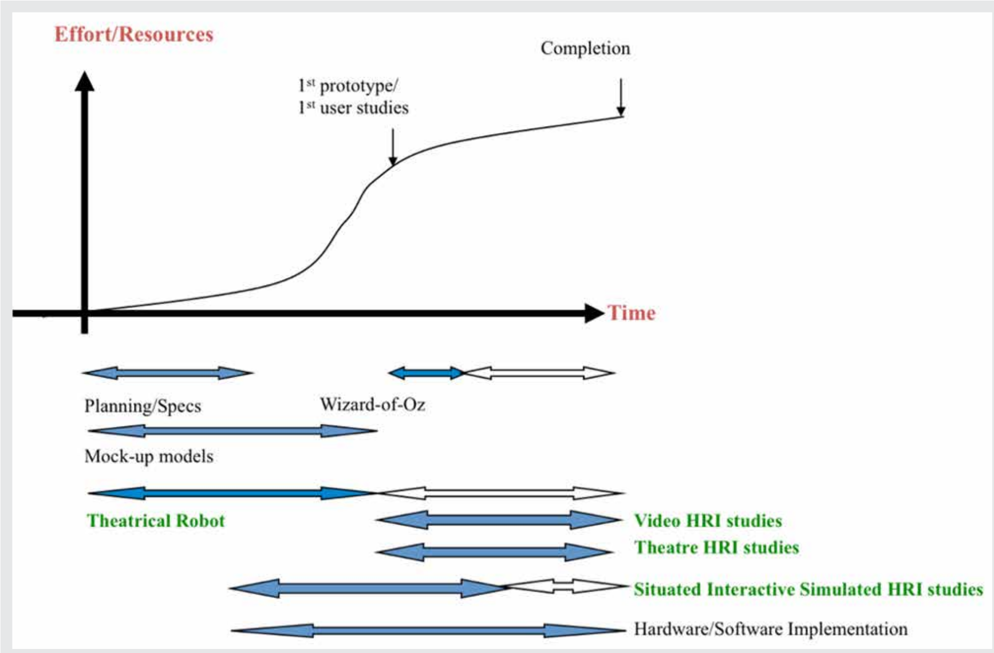



FIGURE 38.2: The iCub (2013) humanoid open course platform, developed as part of the Robotcub project (2013).



FIGURE 38.3: The IROMEC robot which was developed as part of the IROMEC project (2013).

Thus, designing robots for HRI 'properly', i.e. involving users in the design and ensuring that the to be developed robot fulfills its targeted roles and functions and provides positive user experience remains a difficult task (Marti and Bannon, 2009). A number of methods are thus used to gain input and feedback from users before the completion of a fully functioning robot prototype, see Fig. 4. Fig. 5 provides a conceptual comparison of these different prototyping approaches and experimental paradigms.




*Courtesy of Kerstin Dautenhahn. Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms below.*

FIGURE 38.4: Modified from Dautenhahn (2007b), sketching a typical development time line of HRI robots and showing different experimental paradigms. The dark arrows indicate that for those periods the particular experimental method is more useful than during other periods. Note, there are typically several iterations in the development process (not shown in the diagram), since systems may be improved after feedback from user studies with the complete prototype. Also, several releases of different systems may result, based on feedback from deployed robots after a first release to the user/scientific community.

| | Situatedness of interaction | Embodied nature of interaction | Ecological validity | Contingency of interaction | Resource efficiency | Sample size | Realism of interaction/ Outcome-relative fidelity |
|---|---|---|---|---|---|---|---|
| TR | +++ | +++ | + | +++ | +++ | - | - |
| VHRI | + | - | ++ | - | ++ | +++ | - |
| THRI | ++ | - | ++ | - | ++ | +++ | + |
| SISHRI | +++ | - | +++ | +++ | +++ | - | + |
| Live HRI | +++ | +++ | +++ | +++ | - | - | +++ |

Figure 38.5: Conceptual Comparison of Different Experimental Paradigms discussed in this chapter. TR (Theatrical Robot), VHRI (Video-based HRI), THRI (Theatre-based HRI), SISHIR (Situated Interactive Simulated HRI), Live HRI. Resource efficiency means that experiments need to yield relevant results quickly and cheaply (in terms of effort, equipment required, person months etc.). Outcome-relative fidelity means that outcomes of the study must be sufficiently trustworthy and accurate to support potentially costly design decisions taken based on the results (Derbinsky et al. 2013).

Even before a robot prototype exists, in order to support the initial phase of planning and specification of the system, mock-up models might be used, see e.g. Bartneck and Jun 2004. Once a system's main hardware and basic control software has been developed, and safety standards are met, first interaction studies with participants may begin.

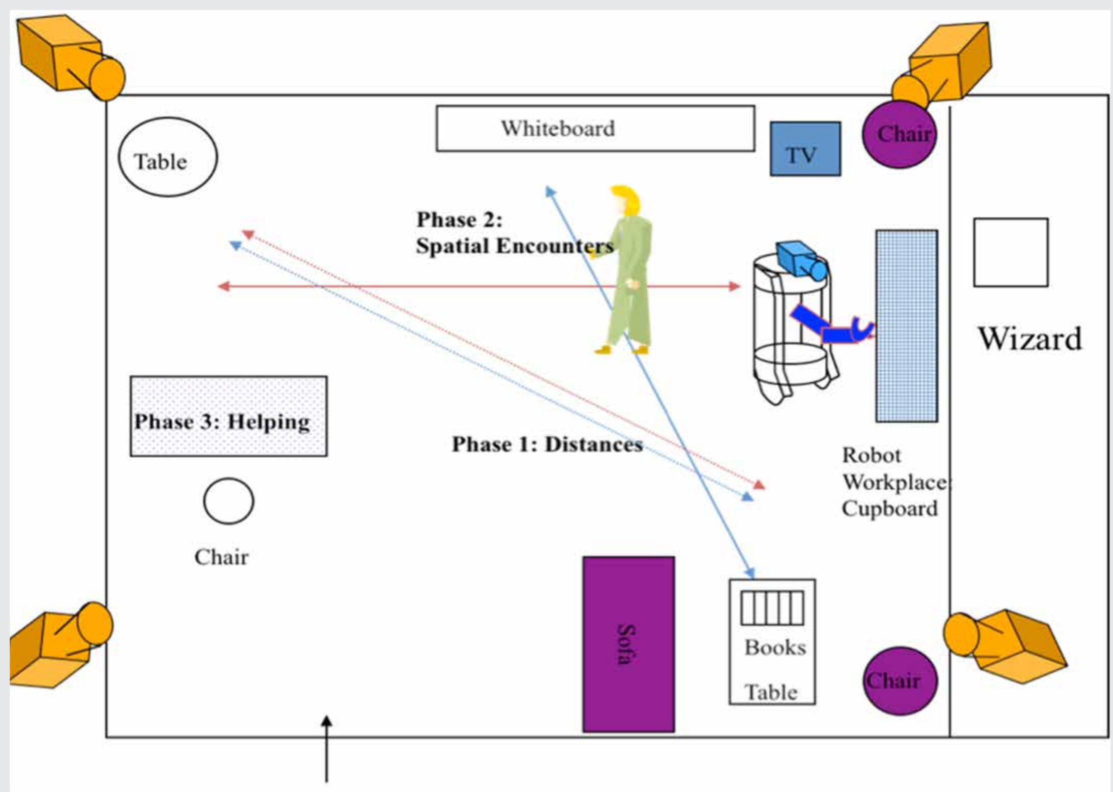
The above mentioned Wizard-of-Oz technique (WoZ) is a popular evaluation technique that originated in HCI (Gould et al, 1983; Dahlback et al., 1993; Maulsby et al,. 1993) and is now widely used in HRI research (Green et al. 2004, Koay et al., 2008; Kim et al., 2012). In order to carry out WoZ studies, a prototype version must be available that can be remotely controlled, unknown to the participants. Thus, WoZ is often used in cases where the robot's hardware has been completed but the robot's sensory, motor or cognitive abilities are still limited. However, having one or two researchers remotely controlling the robot's movements and/or speech can be cognitively demanding and impractical in situations where the goal is that the robot eventually should operate autonomously. For example, in a care, therapy or educational context, remotely controlling a robot require another researcher and/or care staff member to be available (cf. Kim et al., 2013). WoZ can be used for full teleoperation or for partial control, e.g. to simulate the high-level decision-making progress of the robot. See Fig. 6 for an example of an HRI experiment using WoZ.





*Courtesy of Kerstin Dautenhahn. Copyright: CC-Att-SA (Creative Commons Attribution-ShareAlike 3.0 Unported).*

*Courtesy of Kerstin Dautenhahn. Copyright: CC-Att-SA (Creative Commons Attribution-ShareAlike 3.0 Unported).*




*Courtesy of Kerstin Dautenhahn. Copyright: CC-Att-SA (Creative Commons Attribution-ShareAlike 3.0 Unported).*

Figure 38.6: a) Two researchers controlling movement and speech of a robot used in a (simulated) home companion environment (b). 28 subjects interacted with the robot in physical assistance tasks (c), and they also had to negotiate space with the robot (d), e) layout of experimental area for WoZ study. The study was performed in 2004 as part of the EU project COGNIRON. Dautenhahn (2007a), Woods et al. (2007), Koay et al. (2006) provide some results from these human-robot interaction studies using a WoZ approach.

Once WoZ experiments are technically feasible, video-based methods can be applied whereby typically groups of participants are shown videos of the robots interacting with people and their environments. The VHRI (Video-based HRI) methodology has been used successfully in a variety of HRI studies (Walters et al., 2011; Severinson-Eklund, 2011; Koay et al. 2007, 2011; Syrdal et al., 2010; Lohse et al., 2008; Syrdal et al., 2008). Previous studies compared live HRI and video-based HRI and found comparable results in a setting where a robot approached a person (Woods et al., 2006a,b). However, in the scenarios that were used for the comparative study there was little dynamic interaction and co-ordination between the robot's and the person's behaviour. It can be expected that the higher the contingency and co-ordination between human and robot interaction, the less likely VHRI is to simulate live interaction experience (cf. Figure 7).

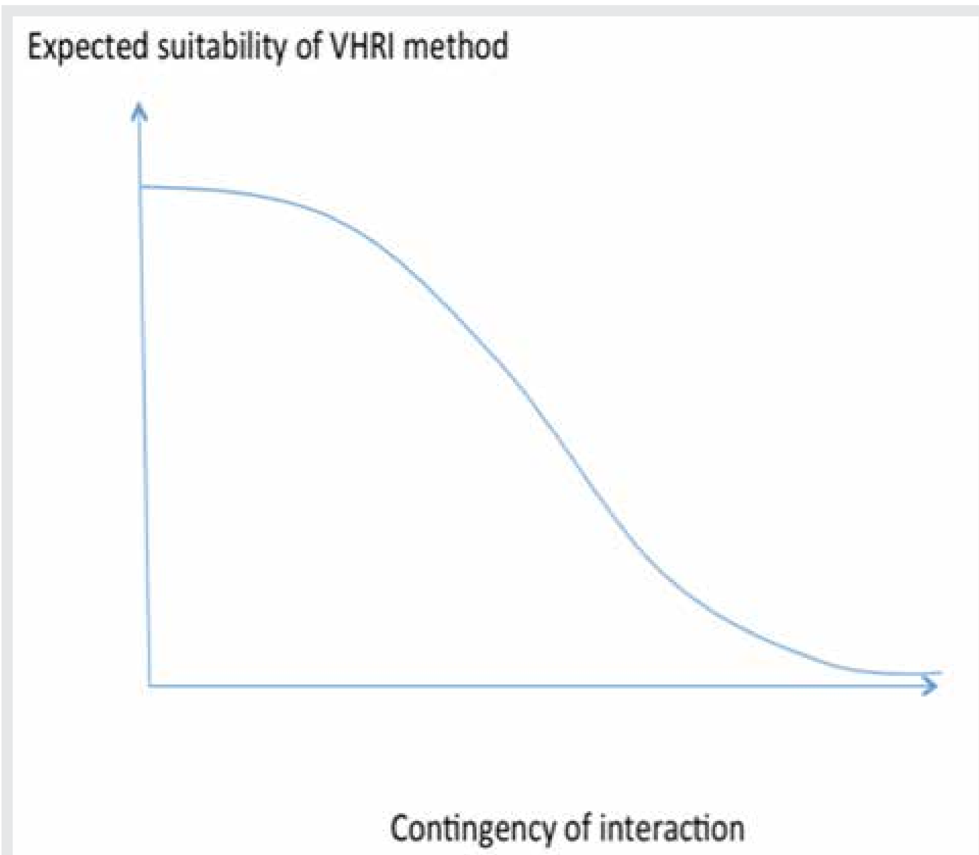





Figure 38.7: Illustration of decrease of suitability of the Video HRI method with increasing contingency of the interaction (e.g. verbal or non-verbal coordination among the robot and the human in interaction).

Another prototyping method that has provided promising results is the Theatrical Robot (TR) method that can be used in instances where a robot is not yet available, but where live human—robot interaction studies are desirable, for example see Fig. 8. The Theatrical Robot describes a person (a professional such as an actor, or mime artist) dressed up as a robot and behaving according to a specific and pre-scripted robotic behaviour repertoire. Thus, the Theatrical Robot can serve as a *life-sized, embodied, simulated robot* that can simulate human-like behaviour and cognition. Robins et al. (2004) have used this method successfully in studies which tried to find out how children with autism react to life-sized robots, and how this reaction depends on whether the robot looks like a person or looks like a robot. The small group of four children studied showed strong initial preferences for the Theatrical Robot in its robotic appearance, compared to the Theatrical Robot showing the same (robotic) behaviour repertoire but dressed as a human being, see example results in Figure 8. Note, in both conditions the 'robot' was trained to not to respond to the children. In the Robins et al. (2004) study a mime artist was used in order to ensure that the TR was able to precisely and reliable control his behaviour during the trials.

The Theatrical Robot paradigm allows us to conduct user studies from an very early phase of planning of the robotic system. Once working prototypes exist the TR method is less likely to be useful since now studies can be run with a 'real' system. However, the TR can also be used as a valuable method on its own, in terms of investigating how people react to other people depending on their appearance, or how people would react to a robot that looks and behaves very human-like. Building robots that truly look and behave like human beings is still a future goal, although Android robots can simulate appearance, they lack human-like movements, behaviour and cognition (MacDorman, Ishiguro, 2006). Thus, the TR can shortcut the extensive development process and allow us to make predictions of how people may react to highly human-like robots.











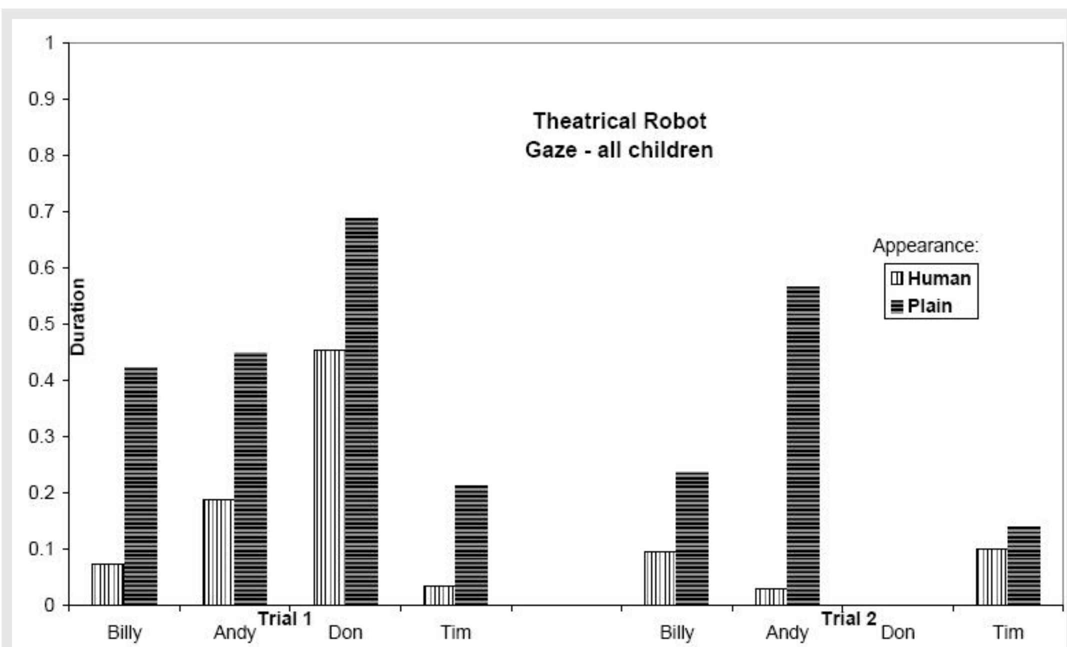
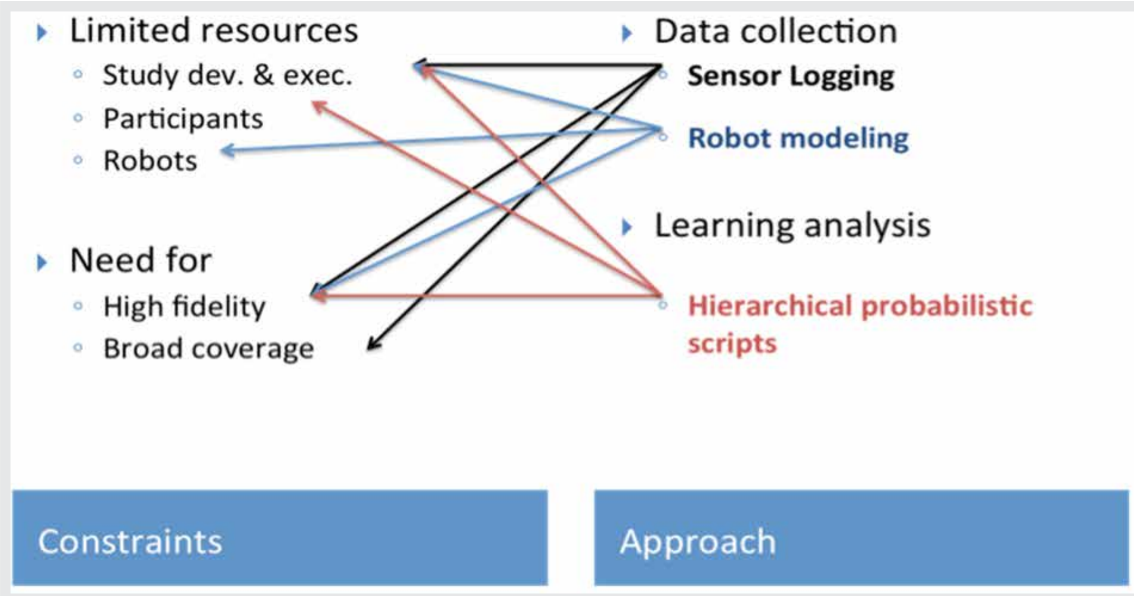




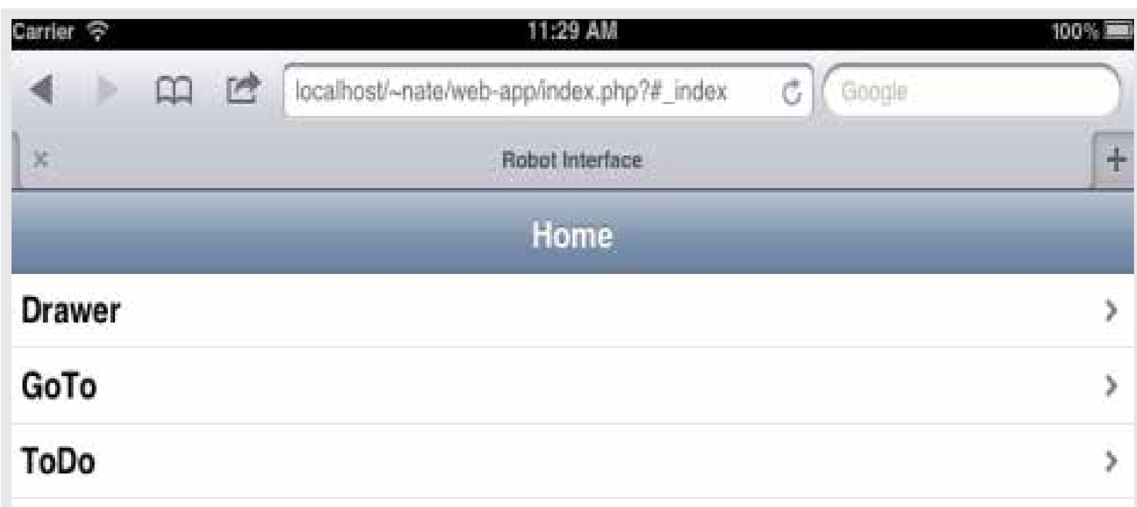

FIGURE 38.8: Using the Theatrical Robot paradigm in a study that investigated children with autism's responses to a human-sized robot either dressed as a robot

(plain appearance a), or as a human person (human appearance),showing identical behaviour in both conditions. In b, c, d responses of three children in both experimental conditions are shown. An example of results showing gaze behaviour of the children towards the TR is shown in e).

In addition to prototyping robots and human-robot interaction, a key problem in many HRI studies is the prototyping of scenarios. For example, in the area of developing home companion robots, researchers study the use of robots for different types of assistance, physical, cognitive and social assistance. This may include helping elderly users at home with physical tasks (e.g. fetch-and-carry), reminding users of appointment, events, or the need to take medicine (the robot as a cognitive prosthetic), or social tasks (encouraging people to socialize, e.g. call a friend or family member or visit a neighbor). Implementing such scenarios presents again a huge developmental effort, in particular when the robot's behaviour should be autonomous, and not fully scripted, but adapt to users' individual preferences and their daily life schedule. One way to prototype a scenario is to combine a WoZ method with robotic theatre performance in front of an audience. The Theatre-based HRI method (THRI) has provided valuable feedback into users' perception of scenarios involving e.g. home companion robots (Syrdal et al, 2011; Chatley et al., 2010). Theatre and drama has been used in Human-Computer Interaction to explore issues of the use of future technologies (see e.g. Iacuccui and Kuuti, 2002; Newell et al., 2006). In the context of HRI, THRI consists of a performance of actors on stage interacting with robots that are WoZ controlled, or semi-autonomously controlled. Subsequent discussions with the audience, and/or questionnaires and interviews are then used to study the audience's perception of the scenarios and the displayed technology. Discussions between the audience and the actors on stage (in character) is typically mediated by a facilitator. This method can reach larger audiences than individual HRI studies would provide, and can thus be very useful to prototype scenarios.




*Courtesy of Kerstin Dautenhahn. Copyright: CC-Att-SA (Creative Commons Attribution-ShareAlike 3.0 Unported).*




*Courtesy of Kerstin Dautenhahn. Copyright: CC-Att-SA (Creative Commons Attribution-ShareAlike 3.0 Unported).*




FIGURE 38.9 A-B-C: a) SISHRI methodological approach (Derbinsky et al., 2013) — situated, real-time HRI with a simulated robot to prototype scenarios, b) example of simulation of interaction shown on the tablet used by the participant. On the left, the homepage of the web application developed for rapid scenarios prototyping is shown. This demo version shows three actions that were implemented (Drawer, GoTo and ToDo): Drawer gave the user the possibility of opening and closing the robot´s drawer. GoTo is used to simulate the time that the robot will take to travel from one position to other (picture on the right), ToDo was introduced to expand the functionality of this prototype, the activities relate to the user, rather than the robot and can be logged in the system (e.g. drinking, eating, etc).

On the right, the functionality GoTo is represented. In this example, the user can send the robot from the kitchen (current robot position) to any other place the user selects from the list (kitchen, couch, desk, drawer). In the picture, the user has chosen the kitchen.

Recently, a new resource efficient method for scenario prototyping has been proposed. A proof-of-concept implementation is described in Derbinsky et al. (2013). Here, an individual user, with the help of a handheld device, goes 'through the motions' of robot home assistance scenarios without an actual physical robot. The tablet computer simulates the robot's actions as embedded in a smart environment. The advantage of this method is that the situatedness of the interaction has been maintained, i.e. the user interacts in a real environment, in real time, with a simulated robot. This method, which can be termed SISHRI (Situated Interactive Simulated HRI) maintains the temporal and spatial aspects and the logical order of action sequences in the scenario, but omits the robot. It allows testing of acceptability and general user experience of complex scenarios, e.g. scenarios used for home assistance without requiring a robot. The system responds based on activities recognized via the sensor network and the input from the user via the user interface. The method is likely to be most useful to prototype complex scenarios before an advanced working prototype is available (see Fig. 9).

*Courtesy of Kerstin Dautenhahn. Copyright: CC-Att-SA (Creative Commons Attribution-ShareAlike 3.0 Unported).*


*Courtesy of Kerstin Dautenhahn. Copyright: CC-Att-SA (Creative Commons Attribution-ShareAlike 3.0 Unported).*


*Courtesy of Kerstin Dautenhahn. Copyright: CC-Att-SA (Creative Commons Attribution-ShareAlike 3.0 Unported).*

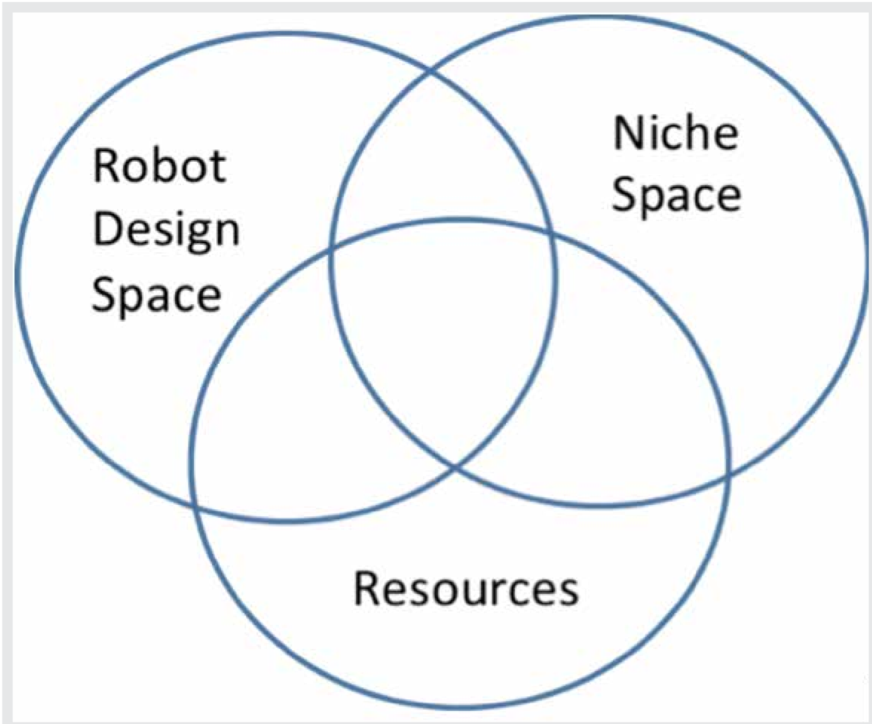





Figure 38.10: Illustrating the design space of robots. Shape and functionality are dependent on their application and use. a) KASPAR, the minimally expressive robot developed at University of Hertfordshire used for HRI studies including robot-assisted therapy for children with autism, b) Roomba (iRobot), a vaccuming cleaning robot shown operating in the University of Hertfordshire Robot House, c) Autom, the weight loss coach. Credit: Intuitive Automata, d) Pleo robot (Ugobe), designed as a 'care-receiving' robot encouraging people to develop a relationship with it e) Robosapien toy robot (WowWee), f) Design space — niche space — resources, see main text for explanation.

The development of any particular HRI study and the methodologies used need to consider the three key constraints shown in Fig. 10. The Robot Design Space comprises all the different possible designs in terms of robot behaviour and appearance. The Niche Space consists of the requirements for the robot and the human-robot interaction as relevant particular scenarios and application areas. The resources (in terms of time, funding, availability of participants etc.) need to be considered when selecting any particular method for HRI studies. Exhaustively exploring the design space is infeasible, so decisions need to be made carefully.

## 38.4 HRI - ABOUT (NOT) ROMANTICIZING ROBOTS

The present reality of robotics research is that robots are far from showing any truly human-like abilities in terms of physical activities, cognition or social abilities (in terms of flexibility, “scaling up” of abilities, “common sense”, graceful degradation of competencies etc.). Nevertheless, in the robotics and HRI literature they are often portrayed as “friends”, “partners”, “co-workers”, etc., all of which are genuinely human terms. These terms are rarely used in an operational sense, and few definitions exist—most often these terms are used without further reflection. Previously, I proposed a more formal definition of *companion robots*, i.e. “A robot companion in a home environment needs to ’do the right things’, i.e. it has to be useful and perform tasks around the house, but it also has to ’do the things right’, i.e. in a manner that is believable and acceptable to humans” (Dautenhahn, 2007a, p. 683).

In contrast to the companion paradigm, where the robot’s key function is to take care of the human’s needs, in the *caretaker paradigm* it is the person’s duty to take care of the ’immature’ robot. In that same article I also argued that due to evolutionarily determined cognitive limits we may be constrained in how many “friends” we may make. When humans form relationships with people, this entails emotional, psychological and physiological investment. We would tend to make a similar investment towards robots, which do not reciprocate this investment. A robot will ’care’ about us as much or as little as the programmers want it to. *Robots are not people; they are machines*. Biological organisms, but not robots, are sentient beings, they are alive, they have an evolutionary and developmental history, they have life-experiences that are shaping their behaviour and their relationships with the environment. In contrast, machines are neither alive nor sentient;

they can express emotions , pretend to ’bond’ with you, but these are simulations, not the real experiences that humans share. The ’emotions’ of a humanoid robot may look human-like but the robot does not feel anything, and the expressions are not based on any experiential understanding. A humanoid robot which looks deeply into your eyes and mutters “I love you” — is running a programme. We may enjoy this interaction, in the way we enjoy role play or immersing ourselves in imaginary worlds, but one needs to be clear about the inherently mechanical nature of the interaction. As Sherry Turkle has pointed out, robots as ’relational artifacts’ that are designed to encourage people to develop a relationship with them, can lead to misunderstandings concerning the authenticity of the interaction (Turkle, 2007). If children grow up with a robot companion as their main friend who the interact with for several hours each day, they will learn that they can just switch it off or lock it into a cupboard whenever it is annoying or challenging them. What concept of friendship will these children develop? Will they develop separate categories, e.g. ’friendship with a robot’, ’friendship with pets’ and ’friendship with people’? Will they apply the same moral and ethical concerns to robots, animals and people? Or will their notion of friendship, shaped by interactions with robots, spill over to the biological world? Similar issues are discussed in terms of children’s possible addiction to computer games and game characters and to what extent these may have a negative impact on their social and moral development. Will people who grow up with a social robot view it as a ’different kind’, regardless of its human or animal likeness? Will social robots become new ontological categories (cf. Kahn et al. 2004; Melson et al. 2009)? At present such questions cannot be answered, they will require long-term studies into how people interact with robots, over years or decades — and such results are difficult to obtain and may be ethically undesirable. However, robotic pets for children and robotic assistants for adults are becoming more and more widespread, so we may get answers to these questions in the future. The answers are unlikely to be ’black and white’ — similar to the question of whether computer games are beneficial for children’s cognitive, academic and social development, where answers are inconclusive (Griffiths, 2002; Kierkegaard, 2008; Dye et al., 2009; Anderson et al., 2010; Jackson et al. 2011).

Humans have been fascinated by autonomous machines throughout history, so the fascination with robots, what they are and what they can be, will stay with us for a long time to come. However, it is advisable to have the discussion on the nature of robots based on facts and evidence, and informed predictions, rather than pursuing a romanticizing fiction.

## 38.5 HRI - THERE IS NO SUCH THING AS 'NATURAL INTERACTION'

A widespread assumption within the field of HRI is that 'good' interaction with a robot must reflect *natural* (human-human) interaction and communication as closely as possible in order to ease people's need to interpret the robot's behaviour. Indeed, people's face-to-face interactions are highly dynamic and multi-modal — involving a variety of gestures, language (content as well as prosody are important), body posture, facial expressions, eye gaze, in some contexts tactile interactions, etc. This has lead to intensive research into how robots can produce and understand gestures, how they can understand when being spoken to and respond correspondingly, how robots can use body posture, eye gaze and other cues to regulate the interaction, and cognitive architectures are being developed to provide robots with natural social behaviour and communicative skills (e.g. Yamaoka et al. 2007; Shimada and Kanda, 2012; Salem, 2012; Mutlu et al., 2012). The ultimate goal inherent in such work is to create human-like robots, which look human-like and behave in a human-like manner. While we discuss below in more detail that the goal of human-like robots needs to be reflected upon critically, the fundamental assumption of the existence of 'natural' human behaviour is also problematic. What is *natural* behaviour to begin with? Is a person behaving naturally in his own home, when playing with his children, talking to his parents, going to a job interview, meeting colleagues, giving a presentation at a conference? The same person behaves differently in different contexts and at different times during their lifetime. Were our hunter-gatherer ancestors behaving naturally when trying to avoid big predators and finding shelter? If 'natural' is meant to be 'biologically realistic' then the argument makes sense — a 'natural gesture' would then be a gesture using a biological motion profile and an arm that is faithfully modeling human arm morphology. Similarly, a natural smile would then try to emulate the complexity of human facial muscles and emotional expressions. However, when moving up from the level of movements and actions to social behaviour, the term 'natural' is less meaningful. To give an example, how polite shall a robot be? Humans show different behaviour and use different expressions in situations where we attend a formal work dinner, or are having a family dinner at home. As humans, we may have many different personal and professional roles in life, e.g. daughter/son, sibling, grandmother, uncle, spouse, employee, employer, committee member, volunteer, etc. We will behave slightly differently in all these different circumstances, from the way we dress, speak, behave, what we say and how we say it, it influences our style of interaction, the manner we use tactile interaction, etc. We can seamlessly switch between these different roles, which are just different aspects of 'who we are' — as expressions of our self or our 'centre of narrative gravity' as it has been phrased by Daniel Dennett. People can deal with such different situations since we continuously re-construct the narratives of our (social) world (Dennett, 1989/91; see also Turner, 1996).

> "*Our fundamental tactic of self-protection, self-control, and self-definition is not building dams or spinning webs, but telling stories - and more particularly concocting and controlling the story we tell others - and ourselves - about who we are.*
>
> *These strings or streams of narrative issue forth as if from a single source - not just in the obvious physical sense of flowing from just one*

*mouth, or one pencil or pen, but in a more subtle sense: their effect on any audience or readers is to encourage them to (try to) posit a unified agent whose words they are, about whom they are: in short, to posit what I call a center of narrative gravity (Dennett, 1989/91).”*

Thus, for humans, behaving ’naturally’ is more than having a given or learnt behaviour repertoire and making rational decisions in any one situation on how to behave. We are ’creating’ these behaviours, reconstructing them, taking into consideration the particular context, interaction histories, etc., we are creating behaviour consistent with our ’narrative self’. For humans, such behaviour can be called ’natural’.

What is ’natural’ behaviour for robots? Where is the notion of ’self’, their ’centre of narrative gravity’? Today’s robots are machines, they may have complex ’experiences’ but these experiences are no different from those of other complex machines. We can program them to behave differently in different contexts, but from their perspective, it does not make any difference whether they behave one way or the other. They are typically not able to relate perceptions of themselves and their environment to a narrative core, they are not re-creating, but rather recalling, experience. Robots do not have a genuine evolutionary history, their bodies and their behaviour (including gestures etc.) have not evolved over many years as an adaptive response to challenges in the environment. For example, the shape of our human arms and hands has very good ’reasons’, it goes back to the design of forelimbs of our vertebrate ancestors, used first for swimming, then as tetrapods for walking and climbing, later bipedal postures freed the hands to grasp and manipulate objects, to use tools, or to communicate via gestures. The design of our arms and hands is not accidental, and is not ’perfect’ either. But our arms and hands embody an evolutionary history of adaptation to different environmental constraints. In contrast, there is no ’natural gesture’ for a robot, in the same way as there is no ’natural’ face or arm for a robot.

To conclude, there appears to be little argument to state that a particular behaviour X is natural for a robot Y. Any behaviour of a robot will be natural or artificial, solely depending on how the humans interacting with the robot perceive it. Thus, naturalness of robot behaviour is in the eyes of the beholder, i.e. the human interacting with or watching the robot; it is not a property of the robot's behaviour itself.

## 38.6 HRI - NEW ROLES

While more and robotic systems can be used in 'the wild' (Sabanovic et al., 2006; Salter et al., 2010) researchers have discussed different roles for such robots.

Previously,I proposed different roles of robots in human society (Dautenhahn, 2003), including:

- a machine operating without human contact;
- a tool in the hands of a human operator;
- a peer as a member of a human—inhabited environment;
- - a robot as a persuasive machine influencing people's views and/ or behaviour (e.g. in a therapeutic context);
- a robot as a social mediator mediating interactions between people;
- a robot as a model social actor.

Dautenhahn et al.(2005) investigated people's opinions on viewing robots as friends, assistants or butlers. Others have discussed similar roles of robots and humans,e.g. humans can assume the role of a supervisor, an operator, a mechanic, a peer, or a bystander (Scholtz, 2003). Goodrich and Schultz (2007) have proposed roles for a robot as a mentor for humans or information consumer whereby a human uses information provided by a robot. Other roles that have been discussed recently are robots as team member in collaborative tasks (Breazeal et al. 2004), robots as learners (Thomaz and

Breazeal, 2008; Calignon et al., 2010; Lohan et al., 2011), and robots as cross-trainers in HRI teaching contexts (Nikolaidis & Shal, 2013). Teaching robots movements, skills and language in interaction and/or by demonstration is a very active area of research (e.g. Argall et al. 2009; Thomaz and Cakmak 2009; Konidaris et al., 2012; Lyon et al. 2012; Nehaniv et al., 2013), however, it remains a challenge on how to teach in natural, unstructured and highly dynamic environments. For humans and some other biological species *social learning* is a powerful tool for learning about the world and each other, to teach and develop culture, and it remains a very interesting challenge for future generations of robots learning in human-inhabited environment. (Nehaniv & Dautenhahn, 2007). Ultimately, robots that can learn flexibly, efficiently, and socially appropriate behaviours that enhance its own skills and performance and is acceptable for humans interacting with the robot, will have to develop suitable levels of social intelligence (Dautenhahn, 1994, 1995, 2007a).

## 38.7 ROBOTS AS SERVICE PROVIDERS

A lot of research in intelligent, autonomous robots has focused on how the robots could provide *services* (assistive or otherwise) that originally people performed. Robots replaced many workers at the factory assembly lines, and more recently robots have been discussed e.g. in the context of providing solutions to care for elderly people in countries with rapidly changing demographics (see Fig. 11). In many scenarios, robots are meant to work alongside people, and to replace some tasks that previously humans performed.

Recently, a number of projects worldwide investigate the use of robots in elder-care in order to allow users to live independently in their homes for as long as possible see e.g. Heylen et al. (2012), Huijnen et al. (2011). Such research poses many technological, ethical and user-related challenges, for examples of such research projects see Fig. 6 for HRI research on home companions in the COGNIRON project (2004-2008), Fig. 12 for research in the LIREC project (2008-

2012), and Fig. 1 for social and empathic home assistance in a smart home as part of the above mentioned ACCOMPANY project. Many such projects use a smart home environment, e.g. the University of Hertfordshire Robot House which is equipped with dozens of sensors. Success in this research domain will depend on acceptability, not only by the primary users of such systems (elderly people) but also by other users (family, friends, neighbours) including formal and informal carers. Thus, taking into consideration the 'human component' is important for such projects. See Amirabdollahian et al. (2013) for a more detailed discussion of the objectives and approaches taken in the ACCOMPANY project.

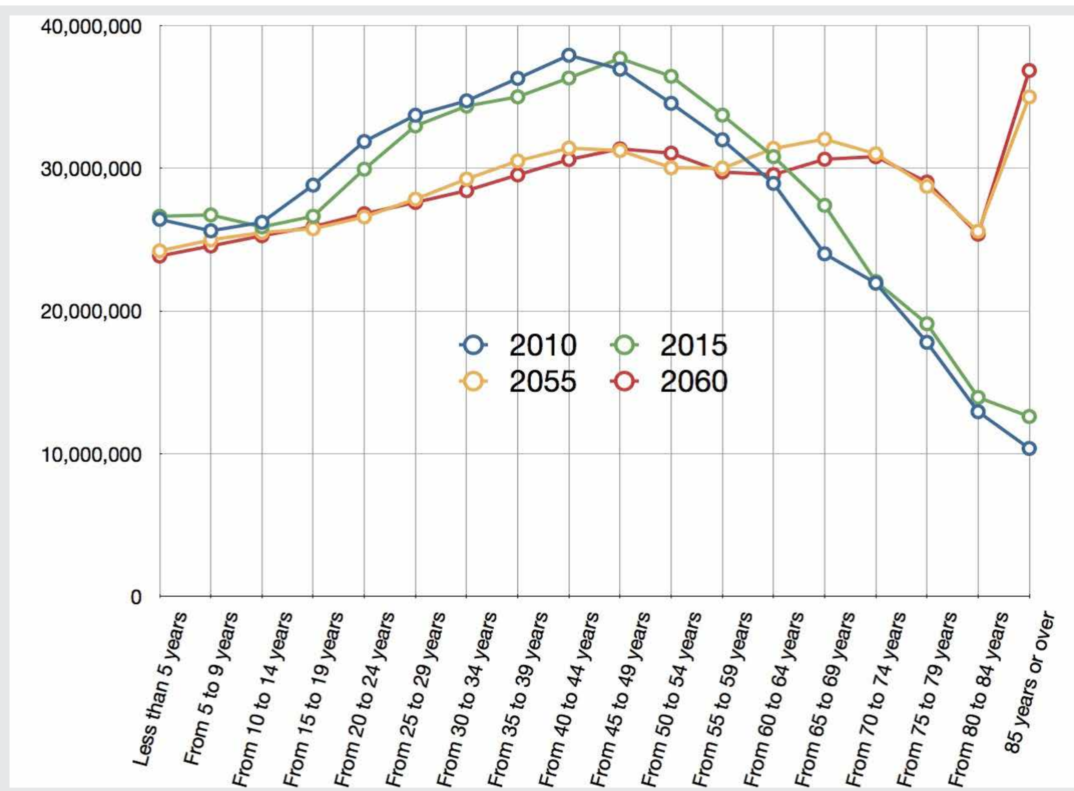





Figure 38.11: Population projections for the 27 Member States, showing an increase of people aged 65 and above from 17.57% to 29.54%, with a decrease of people aged

between 15-64 from 67.01% to 57.42%. Diagram taken with permission from Amirabdollahian et al. (2013).

Note, the domain of robots for elder-care poses many ethical challenges (see e.g. Sharkey and Sharkey, 2011, 2012), and the investigation of these issues is indeed one of the aims of the ACCOMPANY project. In the following I like to provide some personal thoughts on some of these matters. Often robots are envisaged as providing company and social contact, stimulation, motivation, and also facilitating communication among e.g. residents in a care home; see many years of studies with the seal robot PARO (Wada and Shibata, 2007; Shibata et al. 2012). Indeed, care staff has often very little time (typically in the range of a few minutes per day per person), for social contact. So care providers may show a great interest in using robots for social company, and elderly people might welcome such robots as a means to combat their loneliness. However, as I have argued above, interactions with robots are inherently mechanical in nature; robots do not reciprocate love and affection, they can only simulate those. Thus, human beings are and will remain the best experts on providing social contact and company, experiencing and expressing empathy, affection, and mutual understanding. While it is difficult to design robots that can do the more practical tasks that dominate the work day of care staff, e.g. cleaning, feeding, washing elderly people, robots may be designed to fulfill those tasks, potentially freeing up care staff to provide social contact with genuine, meaningful interactions. Unfortunately, it is technically highly challenging to build robots that can actually provide such tasks, although it is an active area of research (cf. the RI-MAN robot and Yamazaki et al., 2012), while it is well within our reach to build robots that provide some basic version of company and social interaction, 'relational artifacts' according to Turkle et al. (2006), that already exist today. If one day robots are able to provide both social and non-social aspects of care, will human care staff become obsolete due to the need of cutting costs in elder-care? Or will robots be used to do the routine work and the time of human carers will be freed to engage with elderly residents in meaningful and emotionally satisfying ways? The latter option would not only be more successful in providing efficient and at the same time humane care, it would also acknowledge our biological roots, emotional needs, and evolutionary history—as a species, our social skills are the one domain where we typically possess our greatest expertise, while our 'technical/mechanical' expertise can be replaced more easily by machines.

*Courtesy of Kerstin Dautenhahn. Copyright: CC-Att-SA (Creative Commons Attribution-ShareAlike 3.0 Unported).*

*Courtesy of Kerstin Dautenhahn. Copyright: CC-Att-SA (Creative Commons Attribution-ShareAlike 3.0 Unported).*





Figure 38.12: The Sunflower robot developed by Dr. Kheng Lee Koay at the University of Hertfordshire. Based on a Pioneer mobile platform (left), a socially interactive and expressive robot was developed for the study of assistance scenarios for a robot companion in a home context.

An example of a robot designed specifically for home assistance is the Sunflower robot illustrated in Figure 12. It consists of a mobile base, a touch-screen user interface and diffuse LED display panels to provide expressive multi-coloured light signals to the user. Other expressive behaviours include sound, base movement, and movements of the robot's neck. The non-verbal expressive behaviours have been inspired by expressive behaviour that dogs display in human-dog interaction in similar scenarios as those used in the Robot House, in collaboration with ELTE in Hungary (Prof. Ádám Miklósi's group). The robot possesses some human-like features (a head, arms) but its overall design is non-humanoid. This design follows our previous research results showing that mechanoid (mechanically-looking) robots are well accepted by users with different individual preferences. The robot's expressive behaviour (light, sound, movements) has been inspired by how dogs interact with their owners (Syrdal et al., 2010; Koay et al., 2013). a) early Sunflower prototype, b,c) Sunflower, d) HRI home assistance scenarios with an early Sunflower prototype in comparison to dog-owner interaction in a comparable scenario, e) (Syrdal et al., 2010). For different expressions of Sunflower see (the picture gallery) and (a video).

## 38.8 ROBOTS AS SOCIAL MEDIATORS

Above we discussed the role of robots as service providers, companions and 'helpers'. A complementary view of robots is to consider their role as *social mediators* — machines that help people to connect with each other. Such robots are not meant to replace or complement humans and their work; instead, their key role is helping people to engage with others. One area where robotic social mediators have been investigated is the domain of robot-assisted therapy (RAT) for children with autism.

Autism is a lifelong developmental disorder characterized by impairments in communication, social interaction and imagination and fantasy (often referred to as the triad of impairments; Wing, 1996) as well as restricted interests and stereotypical behaviours. Autism is a spectrum disorder and we find large individual differences in how autism may manifest itself in a particular child (for diagnostic criteria see DSM IV, 2000). The exact causes of autism are still under investigation, and at present no cure exists. A variety of therapeutic approaches exist, and using robots or other computer technology could complement these existing approaches. The prevalence rate for autism spectrum disorders is often reported as around 1 in 100 but statistical data vary.

While in 1979 Weir and Emanuel had encouraging results with one child with autism using a button box to control a LOGO Turtle from a distance, the use of interactive, social robots as therapeutic tools was first introduced by the present author (Dautenhahn (1999)) as part of the Aurora project (1998, ongoing). Very early in this work the concept of a social mediator for children with autism was investigated, with the aim to encourage interaction between children with autism and other people. The use of robots for therapeutic or diagnostic applications has rapidly grown over the past few years, see recent review articles which show the breadth of this research field and the number of active research groups (Diehl et al., 2012, Scassellati et al. 2012), compared to an earlier review (Dautenhahn & Werry, 2004).

In the earliest work of robots as social mediators for children with autism, Werry et al. (2001) and the present author (Dautenhahn 2003) gave examples of trials with pairs of children who started interacting with each other in a scenario where they had to share an autonomous, mobile robot that they could play with. Work with the humanoid robot Robota (Billard et al. 2006) later showed that the robot could encourage children with autism to interact with each other, as well as a co-present experimenter (Robins et al. 2004; Robins et al. 2005a). Note, the role of a robotic social mediator is not to replace, but to facilitate human contact (Robins et al., 2005a,b, 2006). Similarly, recent work with the minimally expressive humanoid robot KASPAR discusses the robot's role as a salient object that mediates and encourages interaction between the children and co-present adults (Robins et al, 2009; Iacono et al., 2011). Figures 13 to 16 give examples of trials conducted by Dr. Ben Robins where robots have been used as social mediators.

A key future challenge of robots as social mediators is to investigate how robots can adapt in real-time to different users. Francois et al. (2009) provide a proof-of-concept study showing how an AIBO robot can adapt to different interaction styles of children with autism playing with it, see also a recent article by Bekele et al., (2013).





Figure 38.13 A-B: KASPAR as a social mediator for children with autism. Two boys playing an imitation game, one child controls the robot's expressions, the other child has to imitate KASPAR, then the children switch roles.

*Courtesy of Kerstin Dautenhahn. Copyright: CC-Att-SA (Creative Commons Attribution-ShareAlike 3.0 Unported).*

*Courtesy of Kerstin Dautenhahn. Copyright: CC-Att-SA (Creative Commons Attribution-ShareAlike 3.0 Unported).*

Courtesy of Kerstin Dautenhahn. Copyright: CC-Att-SA (Creative Commons Attribution-ShareAlike 3.0 Unported).

Courtesy of Kerstin Dautenhahn. Copyright: CC-Att-SA (Creative Commons Attribution-ShareAlike 3.0 Unported).





Figure 38.14: Two children with autism enjoying the imitation game with KASPAR. One child uses a remote control to make KASPAR produce gestures and body postures; the role of the second child is to imitate KASPAR. After a while the roles are switched.





Figure 38.15 A-B: Sharing with another person (an adult on the left, another child on the right) while playing games with KASPAR.

Courtesy of Kerstin Dautenhahn. Copyright: CC-Att-SA (Creative Commons Attribution-ShareAlike 3.0 Unported).

Courtesy of Kerstin Dautenhahn. Copyright: CC-Att-SA (Creative Commons Attribution-ShareAlike 3.0 Unported).





Figure 38.16 A-B-C-D: Two children with autism enjoying a collaborative game with Robota. The robot is remotely controlled by the experimenter. Robota will only move and adopt a certain posture if both children simultaneously adopt this posture. Findings showed that this provided a strong incentive for the children to coordinate their movements.

The second example area for robots being used as social mediators concerns remote human-human interaction.

While in robotics research touch sensors have been used widely e.g. allowing robots to avoid collisions or to pick up objects, the social dimension of human-robot touch has only recently attracted attention. Humans are born social and tactile creatures. Seeking out contact with the world, including the social world, is key to learning about oneself, the environment, others, and relationships we have with the world. Through tactile interaction we develop cognitive, social and emotional skills, and attachment with others. Tactile interaction is the most basic form of how humans communicate with each other (Hertenstein, 2006). Studies have shown the devastating effects that deprivation of touch in early childhood can have (e.g. Davis, 1999).

Social robots are usually equipped with tactile sensors, in order to encourage play and allow the robot to respond to human touch, e.g. AIBO (Sony), Pleo (Ugobe), PARO (Shibata et al., 2012). Using tactile HRI to support human-human communication over distance illustrates the role a robot could play in order to mediate human contact (Mueller et al., 2005; Lee at al. 2008; The et al., 2008; Papadopoulos et al. 2012a,b).

To illustrate this research direction, Fotios Papadopoulos has investigated how autonomous AIBO robots (Sony) could mediate distant communication between two people engaging in online game activities and interaction scenarios. Here, the long-term goal is to develop robots as social mediators that can assist human-human communication in remote interaction scenarios, in order to support, for example, friends and family members who are temporarily or long term prevented from face-to-face interaction. One study compared how people communicate with each other through a communication system named AiBone involving video communication and interaction with and through an AIBO robot with a setting not involving any robots and using standard computer interfaces instead (Papadopoulos et al., 2012). The experiment involved twenty pairs of participants who communicated using video conference software. Findings showed that participants expressed more social cues when using the robot, and shared more of their game experiences with each other. However, results also show that in terms of efficiency of how to perform the tasks (navigating a maze), users performed better without the robot. These results show a careful balance and trade-off between efficiency of interaction and communication modes, and their social relevance in terms of mediating human-human contact and supporting relationships. A second experiment used a less competitive collaborative game called AIBOStory. Using the remote interactive story-telling system participants could collaboratively create and share common stories through an integrated, autonomous robot companion acting as a social mediator between two remotely located people. Following an initial pilot study, the main experiment studied long-term interactions of 10 pairs of participants using AIBOStory. Results were compared with a condition not involving any physical robot. Results suggests user preferences towards the robot mode, thus supporting the notion that physical robots in the role of social mediators, affording touch-based human-robot interaction and embedded in a remote human-human communication scenario, may improve communication and interaction between people (Papadopoulos, 2012b).

The use of robots as social mediators is different from the approach of considering robots as ’permanent’ tools or companions — a mediator is no longer needed once mediation has been successful. For example, a child who has learnt all it can learn from a robotic mediator will no longer need the robot; a couple being separated for a few months will not need remote communication technology any more once they are reunited. Thus, the ultimate goal of a robotic mediator would be to disappear eventually, after the ’job’ has been done.

## 38.9 HRI - THERE IS A PLACE FOR NON-HUMANOID ROBOTS

It is often assumed as 'given' (e.g. not reflected upon) that the ultimate goal for designers of robots for human-inhabited environments is to develop humanoid robots, i.e. robots with a human-like shape, 2 legs, 2 arms, a head, social behaviour and communication abilities similar to human beings. Different arguments are often provided, some technical, others non-technical:

- humanoid robots would be able to operate machines and work in environments that originally were designed for humans, e.g. the humanoid robot would be able to open our washing machine and use our tool box. This would be in contrasted to robots that require a pre-engineered environment.
- in many applications robots are meant to be used in tasks that require human-like body shapes, e.g. arms to manipulate objects, legs to walk over uneven terrain etc.
- the assumption that humanoid robots would have greater acceptability by people, that they mind 'blend in' better, that people would prefer to interact with them. It is argued that people would be able to more easily predict and respond to the robot's behaviour due to its familiarity with human motion and behaviour, and predictability may contribute to safety.
- the assumption that those robots would fulfill better human-like tasks, e.g. operating machinery and functioning in an environment designed for people, or for the purpose of a robot carrying out human-like tasks, e.g. a companion robots assisting people in their homes or in a hospital our care home

Likewise, in the domain of life-like agents, e.g. virtual characters, a similar tendency towards human-like agents can be found. Previously, I described this tendency as the 'life-like agent hypothesis' (Dautenhahn, 1999):

"Artificial social agents (robotic or software) which are supposed to interact with humans are most successfully designed by imitating life, i.e. making the agents mimic as closely as possible animals, in particular humans. This comprises both 'shallow' approaches focusing on the presentation and believability of the agents, as well as 'deep' architectures which attempt to model faithfully animal cognition and intelligence. Such life-like agents are desirable since

1. The agents are supposed to act on behalf of or in collaboration with humans; they adopt roles and fulfill tasks normally done by humans, thus they require human forms of (social) intelligence.
2. Users prefer to interact ideally with other humans and less ideally with human-like agents. Thus, life-like agents can naturally be integrated in human work and entertainment environment, e.g. as assistants or pets.
3. Life-like agents can serve as models for the scientific investigation of animal behaviour and animal minds". (Dautenhahn, 1999)

Argument (3) presented above easily translates to robotic agent and companions, since these may be used to study human and animal behaviour, cognition and development (MacDorman and Ishiguro, 2006). Clearly, the humanoid robot is an exciting area of research, not only for those researchers interested in the technological aspects but also, importantly, for those interested in developing robots with human-like cognition; the goal would be to develop advanced robots, or to use the robots as tools for the study of human cognition and development (cf. the iCub which exemplifies this work, e.g. Metta et al., 2010; Lyon et al. 2012). When trying to achieve human-like cognition, it is best to choose a humanoid platform, due to the constraints and interdependencies of animal minds and bodies (Pfeifer, 2007). Precursors of this work can be found in Adaptive Behaviour and Artificial Life research using robots as models to understand biological systems (e.g. Webb, 2001; Ijspeert et al., 2005).

However, arguments (1) and (2) are problematic, for the following reasons:

Firstly, while humans have a natural tendency to anthropomorphize the world and to engage even with non-animate objects (such as robots) in a social manner (e.g. Reeves and Nass, 1996; Duffy, 2003), a humanoid shape often evokes expectations concerning the robot's ability, e.g. human-like hands and fingers suggest that the robot is able to manipulate objects in the same way humans can, a head with eyes suggests that the robot has advanced sensory abilities e.g. vision, a robot that produces speech is expected also to understand when spoken to. More generally, a human-like form and human-like behaviour is associated with human-level intelligence and general knowledge, as well as human-like social, communicative and empathic understanding. Due to limitations both in robotics technology and in our understanding of how to create human-like levels of intelligence and cognition, in interaction with a robot people quickly realize the robot's limitations, which can cause frustration and disappointment.

Secondly, if a non-humanoid shape can fulfill the robot's envisaged function, then this may be the most efficient as well as the most acceptable form. For example, the autonomous vacuum cleaning robot Roomba (iRobot) has been well accepted by users as an autonomous, but clearly non-humanoid robot. Some users may attribute personality to it, but the functional shape of the robot clearly signifies its robotic nature, and indeed few owners have been shown to treat the robot as a social being (Sung et al., 2007, 2008). Thus, rather than trying to use a humanoid robot operating a vacuum cleaner in a human-like manner (which is very hard to implement), an alternative efficient and acceptable solution has been found. Similarly, the ironing robot built by Siemens (Dressman) does not try to replicate the way humans iron a shirt but finds an alternative, technologically simpler solution.

Building humanoids which operate and behave in a human-like manner is technologically highly challenging and costly in terms of time and effort required, and it is unclear when such human-likeness may be achieved (if ever)in future. But even if such robots were around, would we want them to replace e.g. the Roomba? The current tendency to focus on humanoid robots in HRI and robotics may be driven by scientific curiosity, but it is advisable to consider the whole design space of robots, and how the robot's design may be very suitable for particular tasks or application areas. Non-humanoid, often special purpose machines, such as the Roomba, may provide cheap and robust solutions to real-life needs, i.e. to get the floor cleaned, in particular for tasks that involve little human contact. For tasks that do involve a significant amount of human-robot interaction, some humanoid characteristics may add to the robot's acceptance and success as an interactive machine, and may thus be justified better. Note, the design space of robots is huge, and 'humanoid' does not necessarily mean 'as closely as possible resembling a human'. A humanoid robot such as Autom (2013), designed as a weight loss coach has clearly human-like features, but very simplified features, more reminiscent of a cartoon-design. On the other end of the spectrum towards human-like appearance we find the androids developed by Hiroshi Ishiguro and his team (http://www.geminoid.jp/en/index.html), or David Hanson's robots (2013). However, in android technology the limitations are clearly visible in terms of producing human-like motor control, cognition and interactive skills. Androids have been proposed, though, as tools to investigate human cognition (MacDorman and Ishiguro, 2006).

Thus, social robots do not necessarily need to 'be like us'; they do not need to behave or look like us, but they need to do their jobs well, integrate into our human culture and provide an acceptable, enjoyable and safe interaction experience.

## 38.10 HRI - BEING SAFE

Human safety is a key requirement for robots to perform useful tasks alongside humans in a home environment, an office, etc. In such circumstances, the widely used solutions towards robot safety in industry (e.g. BARA, 2012) are not acceptable (e.g. warning sounds, flashing lights, etc.) or they may not be feasible in the particular environment (e.g. use of enclosures, safety guards etc.). Safety in human-robot interaction in its most basic form shall avoid any physical harm to a human being due to collisions with a robot or part of a robot etc. New developments on the robot's technical features (e.g. reliability, control, sensors) and materials (soft, lightweight etc.) can contribute to human-robot safety (Pervez & Ryu, 2008). In situations where physical human-interaction is involved different strategies can be adopted and metrics developed; compare a review in (De Santis et al.,2008) that identifies different approaches for human-robot safety ranging from design, sensors, software, planning, biomimetics to control solutions to human-robot safety. Research in this domain concerns many different aspects, e.g. the analysis and design of safety aspects, the design of safety for robots via the development of specific mechanical and actuator systems or by exploiting new materials, design of low and medium-level controllers for safe compliance via direct force compliance, and the development of high-level cognition, control and decision-making aspects (Herrmann and Melhuish, 2010).

However, even non-harmful interactions may not be perceived as comfortable (e.g. a robot invading a user's personal space by approaching too close). Thus, we can consider objective parameters of physical safety, as well as subjective parameters of *perceived safety*. The latter is likely to change in long-term interactions when a user gets used to interactions with the robot and understands better its functionalities and limitations, which allows the user to make better predictions about the robot's behaviour. Little research has investigated the use of social cues to enhance the safety of human-robot interactions. Research on safety in human-robot interaction usually focuses on technical requirements for safety, rather than addressing possible human behavioural and social mechanisms. However, humans are able to deal with other people even in potentially dangerous situations (e.g. when on a collision path while walking along a hallway) by utilizing a number of communicative verbal as well as non-verbal coordination mechanisms. There are two main aspects to the use of social cues for enhancing safety with robots: 1) The robot can express social cues and show behaviour which intuitively informs the user that a potentially hazardous action by the robot is imminent or under way. In this case it would be up to the person to take the initiative to modify his/her behaviour to ensure safe interaction with the robot. 2) Alternatively, the robot can actively monitor the user's activities (and/or use information from its interaction history with the user to make predictions about the user's behaviour and activities), and modify its own actions accordingly to avoid unsafe interactions. In the latter case the robot takes the initiative and tries to regulate the interactions with the user in a safe manner. Point 2) above is significantly more technically demanding of robot control and sensor systems, but both approaches have the potential to facilitate safe working of a robot in a human-oriented environment. A combination of both approaches, i.e. human and robot both being 'safety-aware' and collaboratively trying to avoid unsafe situations by mutually being attentive to and adapting to each other's current or predicted actions would be the more 'natural' solution, similar to how people coordinate their actions. However, it would require sophisticated perceptual and predictive abilities of the robot, in dynamic and naturalistic environments with complex tasks.

. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .

Note, humanoid robots are not necessarily safer than other robots as implied in the following statement:

*"They can move around our buildings, they can increasingly use the same tools as us, and perhaps most importantly they have the potential to move in a way that naturally makes sense to us - which makes them safer to be around." (http://www.therobotstudio.com/humanoid-robots.html)*

While the above statement may appear intuitive to non-roboticists, a human-like shape does not necessarily help in predicting the behaviour of a robot. When encountering a human we can make fairly good predictions of their maximum speed or strength, even when meeting an athlete. If we make the same predications for human-like robots we may be fundamentally wrong, and engage in behaviour that may result e.g. in injuries for people. Underestimating the weight of a robot or the behaviour of an industrial-strength manipulator arm is clearly not safe, regardless of how human-like they may appear. Thus, safe human-robot interaction needs to be studied carefully. In many cases a non-human like machine, which people have little prior expectations of, will make people act in an instinctively cautious manner around the machine, similar to the caution people apply when encountering unknown and potentially dangerous situations. Thus, for 'first encounters', or application areas where people will meet a particular robot only briefly, non-humanoid machines may have advantages over humanoid robots. Non-humanoid robots decrease the expectations in terms of the skills people attribute to them, and they may elicit cautious behaviour in people who will carefully assess the robot's abilities and how one can safely interact with it, rather than assuming that it 'naturally' has human-like abilities and is safe to interact with.

## 38.11 CONCLUSION. HRI - WHAT ROBOTS ARE TODAY

Social robots are a special kind of (embodied) interactive artifact (see Kahn et al., 2004; Melson et al., 2009) that may afford new types of interactions with people, and new roles they may adopt in society may emerge (see Dautenhahn, 2003). People's relationships with such robots will cover a range from "funny toy" to "long-term companion". Future robots may look and behave very differently from how they do today, and we might develop relationships with them and invent usages for them that we cannot envisage at present. Human culture is changing, too, and people's attitudes towards social robots is likely to change the more prevalent and complex robots become. Elder-care robots that are currently under investigation will probably only be mass deployed when today's young people have reached retirement age—a generation used to electronic devices, the internet and World-Wide-Web, gadgets and social networking on an unprecedented scale. They won't be 'naive users'. But even today's participants in HRI studies are not "naive" in a strict sense—they come with particular attitudes towards technology in general and often robots in particular, even when they have never encountered one face-to-face. People tend to anthropomorphize the world around them, and they react socially even to non-humanoid-looking technology. People are also social animals, and they interpret and interact with the animate and inanimate world around them in social terms (Dautenhahn, 2007). They may respond to robots with some biological reactions typically shown towards humans, but this reaction may be influenced by top-down mechanisms of their beliefs about the system (Shen et al. 2011). Future machines may capitalize on these bottom-up (biological) and top-down (psychological) processes and we may create machines that people may develop special relationships with. HRI is a moving target, and so, as HRI researchers, we need to keep moving, too—being flexible and open-minded about the very foundations of our domain and the nature of robots, and being open-minded towards creative solutions to robot design and methodological challenges. Social robots of the future might be different creatures, complex synthetic entities, but they may have unexpected properties and they may even surprise us and make us behave in surprising ways. As a research community we work towards a new science of HRI that can shape these developments for the benefit of us as individuals and our society.

## 38.12 ACKNOWLEDGEMENTS

I would like to thank Joe Saunders, Michael L. Walters and Chrystopher Nehaniv for helpful comments on the manuscript. I would also like to thank the excellent research team in the Adaptive Systems research group at the University of Hertfordshire who created some of the research work cited in this article which has greatly shaped and changed my ideas on social robots and human-robot interaction over the past 13 years.

## 38.13 REFERENCES

*Amirabdollahian, F., Akker, R.op den, Bedaf, S., Bormann, R., Draper, H.R, Gelderblom, G. J., Ruiz, C. Gutierrez, Hewson, D., Iacono, I., Koay, K. L., Krose, B., Marti, P., Prevot-Huille, H., Reiser, U., Sorell, T. and Dautenhahn, K. (2013): Acceptable robotiCs COMPanions for AgeiNg Years - Multidimensional Aspects of Human-System Interactions. In: Proceeding of the 6th International Conference on Human System Interaction HSI´2013 June 6-8, 2013, Sopot, Poland.*

*Anderson, Craig A., Ihori, Nobuko, Bushman, Brad J., Rothstein, Hannah R., Shibuya, Akiko, Swing, Edward L., Sakamoto, Akira and Saleem, Muniba (2010): Violent video game effects on aggression, empathy, and prosocial behaviour in Eastern and Western countries. In Psychological Bulletin, 136 (2) pp. 151-173*

*Argall, Brenna D., Chernova, Sonia, Veloso, Manuela and Browning, Brett (2009): A Survey of Robot Learning from Demonstration. In Robotics and Autonomous Systems, 57 (5) pp. 469-483*

*Asada, Minoru, Hosoda, Koh, Kuniyoshi, Yasuo, Ishiguro, Hiroshi, Inui, Toshio, Yoshikawa, Yuichiro, Ogino, Masaki and Yoshida, Chisato (2009): Cognitive Developmental Robotics: A Survey. In IEEE Transactions on Autonomous Mental Development, 1 (1) pp. 12-34*

*Association, American Psychiatric (2000): The Diagnostic and Statistical Manual of Mental*

*Disorders. Arlington, USA, American Psychiatric Publishing*

Association, *The British Automation and Robot (2013). A practical guide to machine safety application, legislation and standards. Retrieved 30 May 2013 from Bara.org.uk: http://www.bara.org.uk/info/safety/A_Practical_Guide_to_Machine_Safety_Application.pdf*

*Aurora project. URL: http://www.aurora-project.com/. Last accessed 16 April 2013.*

*BARA (The British Automation and Robot Association). A practical guide to machine safety application, legislation and standards. URL: http://www.bara.org.uk/info/info_safety.html. Accessed December, 2012.*

Bartneck, *Christoph and* Hu, *Jun (2004): Rapid Prototyping for Interactive Robots. In:* Proceedings of the 8th Conference on Intelligent Autonomous Systems IAS-8 *2004, Amsterdam, Netherlands. pp. 136-145*

Bekele, *Esubalew T.,* Lahiri, *Utama,* Swanson, *Amy R.,* Crittendon, *Julie A.,* Warren, *Zachary E. and* Sarkar, *Nilanjan (2013): A Step Towards Developing Adaptive Robot-Mediated Intervention Architecture (ARIA) for Children With Autism. In* IEEE Transactions on Neural Systems and Rehabilitation Engineering, *22 (2) pp. 289-299*

Billard, *Aude,* Robins, *Ben,* Nadel, *Jacqueline and* Dautenhahn, *Kerstin (2006): Building Robota, a Mini-Humanoid Robot for the Rehabilitation of Children with Autism. In* RESNA Assistive Technology Journal, *19 (1) pp. 37-49*

Breazeal, *Cynthia,* Brooks, *Andrew G.,* Gray, *Jesse,* Hoffman, *Guy,* Kidd, *Cory D.,* Lee, *Hans,* Lieberman, *Jeff,* Lockerd, *Andrea and* Chilongo, *David (2004): Tutelage and Collaboration for Humanoid Robots. In* International Journal of Humanoid Robot, *1 (2) pp. 315-348*

Calinon, *Sylvain,* D'halluin, *Florent,* Sauser, *Eric,* Caldwell, *Darwin and* Billard, *Aude (2010): Learning and reproduction of gestures by imitation: An approach based on Hidden Markov Model and Gaussian Mixture Regression. In* IEEE Robotics and Automation Magazine, *17 (2) pp. 44-54*

Cangelosi, *Angelo,* Metta, *Giorgio,* Sagerer, *Gerhard,* Nolfi, *Stefano,* Nehaniv, *Chrystopher,* Fischer, *Kerstin,* Tani, *Jun,* Belpaeme, *Tony,* Sandini, *Giulio,* Fadiga, *Luciano,* Wrede, *Britta,* Rohlfing, *Katharina,* Tuci, *Elio,* Dautenhahn, *Kerstin,* Saunders, *Joe and* Zeschel, *Arne (2010): Integration of Action and Language Knowledge: A Roadmap for Developmental Robotics. In* IEEE Transactions on Autonomous Mental Development, *2 (3) pp. 167-195*

Chatley, *Amiy R.,* Dautenhahn, *Kerstin,* Walters, *Mick L.,* Syrdal, *Dag S. and* Christianson, *Bruce (2010): Theatre as a Discussion Tool in Human-Robot Interaction Experiments - A Pilot Study. In: Conference on Advances in Computer-Human Interactions ACHI 2010 February 10-16, 2010, St. Maarten, Netherlands Antilles. pp. 73-78*

*COGNIRON. URL: http://www.cogniron.org/final/Home.php. Last accessed 11 April 2013*

*Cuijpers, Raymond H., Bruna, Maarten T., Ham, Jaap R. C. and Torta, Elena (2011): Attitude towards Robots Depends on Interaction But Not on Anticipatory Behaviour. In: Mutlu, Bilge, Bartneck, Christoph, Ham, Jaap, Evers, Vanessa and Kanda, Takayuki (eds.). "Social Robotics Lecture Notes in Computer Science". Berlin, Germany: pp. 163-172*

*Dahlback, Nils, Jonsson, Arne and Ahrenberg, Lars (1993): Wizard of Oz Studies -- Why and How. In: Gray, Wayne D., Hefley, William and Murray, Dianne (eds.) International Workshop on Intelligent User Interfaces 1993 January 4-7, 1993, Orlando, Florida, USA. pp. 193-200*

*Dautenhahn, Kerstin (1999): Robots as Social Actors: AURORA and The Case of Autism. In: Proceedings Third Cognitive Technology Conference CT99 August, 1999, San Francisco, USA.*

*Dautenhahn, Kerstin (2007b): Methodology and Themes of Human-Robot Interaction: A Growing Research Field. In International Journal of Advanced Robotic Systems, 4 (1) pp. 103-108*

*Dautenhahn, Kerstin (2003): Roles and functions of robots in human society: implications from research in autism therapy. In Robotica, 21 (4) pp. 443-452*

*Dautenhahn, Kerstin (1995): Getting to know each other-Artificial social intelligence for autonomous robots. In Robotics and Autonomous Systems, 16 (2) pp. 333-356*

*Dautenhahn, Kerstin (1994): Trying to Imitate - a Step Towards Releasing Robots from Social Isolation. In: Gaussier, Philippe and Nicoud, Jean-Daniel (eds.) Proceedings From Perception to Action Conference September 7-9, 1994, Lausanne, Switzerland. pp. 290-301*

*Dautenhahn, Kerstin (2007a): Socially intelligent robots: dimensions of human - robot interaction. In Philosophical Transactions of the Royal Society B: Biological Sciences, 362 (1480) pp. 679-704*

*Dautenhahn, Kerstin and Robins, Ben (2006): The role of the experimenter in HRI research - a case study evaluation of children with autism interacting with a robotic toy. In: The 15th IEEE International Symposium on Robot and Human Interactive Communication RO-MAN 2006 September 6-8, 2006, Hatfield, United Kingdom. pp. 646-651*

*Dautenhahn, Kerstin and Werry, Iain (2004): Towards Interactive Robots in Autism Thera-*

*py: Background, Motivation and Challenges. In Pragmatics and Cognition, 12 (1) pp. 1-35*

Dautenhahn, Kerstin, Woods, Sarah, Kaouri, Christina, Walters, Michael L., Koay, Kheng Lee and Werry, Iain *(2005): What is a Robot Companion - Friend, Assistant or Butler. In: IEEE IRS/RSJ International Conference on Intelligent Robots and Systems August 2-6, 2005, Edmonton, Canada. pp. 1488-1493*

Davis, Phyllis K. *(1999): The Power of Touch - The Basis for Survival, Health, Intimacy, and Emotional Well-Being. Carlsbad,California USA, Hay House Inc*

Dennett, Daniel C. *(1989): The origins of selves. In Cogito, 3 (3) pp. 163-173*

Derbinsky, Nate, Ho, Wan Ching, Duque, Ismael, Saunders, Joe and Dautenhahn, Kerstin *(2013): Resource-Efficient Methods for Feasibility Studies of Scenarios for Long-Term HRI Studies. In:* Miller, Leslie *(ed.) February 24- March 1, 2013, Nice,France. pp. 95-101*

Dickerson, Paul, Robins, Ben and Dautenhahn, Kerstin *(2013): Where the action is: A conversation analytic perspective on interaction between a humanoid robot, a co-present adult and a child with an ASD. In Interaction Studies, 14 (2) pp. 296-316*

Diehl, Joshua J., Schmitt, Lauren M., Villano, Michael and Crowell, Charles R. *(2012): The clinical use of robots for individuals with Autism Spectrum Disorders: A critical review. In Research in Autism Spectrum Disorders, 6 (1) pp. 249-262*

Duffy, Brian R. *(2003): Anthropomorphism and the social robot. In Robotics and Autonomous Systems, 42 (3) pp. 177-190*

Dye, Matthew W.G., Green, Shawn C. and Bavelier, Daphne *(2009): Increasing speed of processing with action video games. In Current Directions in Psychological Science, 18 (6) pp. 321-326*

François, Dorothée, Polani, Daniel and Dautenhahn, Kerstin *(2008): Towards Socially Adaptive Robots: A Novel Method for Real Time Recognition of Human-Robot Interaction Styles. In: 8th IEEE-RAS International Conference on Humanoid Robots Humanoids 2008 December 1-3, 2008, Daejeon, Korea. pp. 353-359*

Goodrich, Michael A. and Schultz, Alan C. *(2007): Human-Robot Interaction: A Survey. In Foundations and Trends in Human-Computer Interaction, 1 (3) pp. 203-275*

Gould, John D., Conti, John and Hovanyecz, Todd *(1983): Composing letters with a simulated listening typewriter. In Communications of the ACM, 26 (4) pp. 295-308*

Green, Anders, Hüttenrauch, Helge and Eklundh, Kerstin Severinson *(2004): Applying the*

*Wizard of Oz Framework to Cooperative Service discovery and Configuration. In: ROMAN 2004 13th IEEE International Workshop on Robot and Human Interactive Communication September 20-22, 2004, Kurashiki, Okayama Japan. pp. 575-580*

*Griffiths, Mark (2002): The educational benefits of video games. In Education and Health, 20 (3) pp. 47-51*

*Harvey, Inman, Paolo, Ezequiel Di, Wood, Rachel, Quinn, Matt and Tuci, Elio (2005): Evolutionary Robotics: A New Scientific Tool for Studying Cognition. In Artificial Life, 11 (1) pp. 79-98*

*Heinzmann, Jochen and Zelinsky, Alexander (2003): Quantitative safety guarantees for physical human-robot interaction. In The International Journal of Robotics Research, 22 (7) pp. 479-504*

*Herrmann, Guido and Melhuish, Chris (2010): Towards safety in human-robot interaction. In International Journal of Social robotics, 2 (3) pp. 217-219*

*Hertenstein, Matthew J., Verkamp, Julie M., Kerestes, Alyssa M. and Holmes, Rachel M. (2006): The communicative functions of touch in humans, non-human primates, and rats: A review and synthesis of the empirical research. In Genetic, Social and General Psychology Monographs, 132 (1) pp. 5-94*

*Heylen, Dirk, Dijk, Betsy van and Nijholt, Anton (2012): Robotic rabbit companions: Amusing or a nuisance. In Journal of Multimodal User Interfaces, 5 (1) pp. 53-59*

*Huijnen, Claire, Badii, Atta, Heuvel, Herjan van den, Caleb-Solly, Praminda and Thiemert, Daniel (2011): "Maybe It Becomes a Buddy, But Do Not Call It a Robot" – Seamless Cooperation between Companion Robotics and Smart Homes. In: Keyson, David V., Maher, Mary Lou, Streitz, Norbert, Cheok, Adrian D., Augusto, Juan C., Wichert, Reiner, Englebienne, Gwenn, Aghajan, Hamid K. and Kröse, Ben J. A. (eds.). "Ambient Intelligence Lecture Notes in Computer Science". Berlin, Germany: Springer Linkpp. 324-329*

*Hüttenrauch, Helge, A.Topp, Elin and Eklundh, Kerstin Severinson (2009): The Art of Gate-Crashing Bringing HRI into users' homes. In Interaction Studies, 10 (3) pp. 274-297*

*Iacono, Iolanda, Lehmann, Hagen, Marti, Patrizia, Robins, Ben and Dautenhahn, Kerstin (2011): Robots as social mediators for children with autism - A preliminary analysis comparing two different robotic platforms. In: IEEE ICDL - EPIROB 2011, first Joint IEEE International Conference on Development and Learning and on Epigenetic Robotics August 24-27, 2011, Frankfurt, Germany. pp. 1-6*

*Iacucci, Giulio and Kuutti, Kari (2002): Everyday Life as a Stage in Creating and Performing*

*Scenarios for Wireless Devices. In Personal and Ubiquitous Computing, 6 (4) pp. 299-306*

Ijspeert, Auke J., Crespi, Alessandro and Cabelguen, Jean-Marie (2005): *Simulation and Robotics Studies of Salamander Locomotion. Applying Neurobiological Principles to the Control of Locomotion in Robots. In Neuroinformatics, 3 (3) pp. 171-196*

Jackson, Linda A., Eye, Alexander von, Fitzgerald, Hiram E., Witt, Edward A. and Zhao, Yong (2011): *Internet use, videogame playing and cell phone use as predictors of children's body mass index (BMI), body weight, academic performance, and social and overall self-esteem. In Computers in Human Behavior, 27 (1) pp. 599-604*

Jr, Peter H. Kahn,, Friedman, Batya, Perez-Granados, Deanne R. and Freier, Nathan G. (2004): *Robotic pets in the lives of preschool children. In: CHI 04 Extended Abstracts on Human Factors in Computing Systems April 24-29, 2004, Vienna, Austria. pp. 1449-1452*

Kanda, Takayuki, Sato, Rumi, Saiwaki, Naoki and Ishiguro, Hiroshi (2007): *A two-month Field Trial in an Elementary School for Long-term Human-robot Interaction. In IEEE Transactions on Robotics, 23 (5) pp. 962-971*

Kanda, Takayuki, Shiomi, Masahiro, Miyashita, Zenta, Ishiguro, Hiroshi and Hagita, Norihiro (2010): *A Communication Robot in a Shopping Mall. In IEEE Transactions on Robotics, 26 (5) pp. 897-913*

Kidd, Cory D. and Breazeal, Cynthia (2008): *Robots at home: Understanding long-term human-robot interaction. In: Proceedings 2008 IEEE/RSJ International Conference on Intelligent Robots and Systems September 22-26, 2008, Nice,France. pp. 3230-3235*

Kierkegaard, Patrick (2008): *Video games and aggression. In International Journal of Liability and Scientific Enquiry, 1 (4) pp. 411-417*

Kim, Elizabeth S., Berkovits, Lauren D., Bernier, Emily P., Leyzberg, Dan, Shic, Frederick, Paul, Rhea and Scassellati, Brian (2012): *Social Robots as Embedded Reinforcers of Social Behavior in Children with Autism. In Journal of Autism and Developmental Disorders, 43 (5) pp. 1038-1049*

Koay, K. L., Dautenhahn, K., Woods, S. N. and Walters, M. L. (2006): *Empirical results from using a comfort level device in human-robot interaction studies. In: Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction 2006. pp. 194-201*

Koay, K. L., Lakatos, G., Syrdal, D.S., Gácsi, M., Bereczky, B., Dautenhahn, Kerstin, Miklosi, A. and Walters, M. L. (2013): *Hey! There is someone at your door. A Hearing Robot using Visual Communication Signals of Hearing Dogs to Communicate Intent. In: IEEE*

ALIFE 2013 April 16-19, 2013, Singapore,Singapore.

Koay, Kheng Lee, Syrdal, Dag Sverre, Walters, Michael L. and Dautenhahn, Kerstin (2009): Five Weeks in the Robot House -- Exploratory Human-Robot Interaction Trials in a Domestic Setting. In: Proceedings of the 2009 International Conference on Advances in Computer-Human Interactions 2009. pp. 219-226

Koay, Kheng Lee, Syrdal, Dag Sverre, Walters, Michael L. and Dautenhahn, Kerstin (2009): A User Study on Visualization of Agent Migration between Two Companion Robots. In: 13th International Conference on Human-Computer Interaction HCII 2009 July 19-24, 2009, San Diego, USA.

Koay, K. L., Syrdal, D. S., Dautenhahn, K., Arent, K., Małek, Ł and Kreczmer, B. (2011): Companion Migration – Initial Participants' Feedback from a Video-Based Prototyping Study. In: Wang, Xiangyu (ed.). "Mixed Reality and Human-Robot Interaction". Dordrecht, Netherlands: Springer Linkpp. 133-151

Konidaris, George, Kuindersma, Scott, Grupen, Roderic and Barto, Andrew (2012): Robot learning from demonstration by constructing skill trees. In The International Journal of Robotics Research, 31 (3) pp. 360-375

Lee, Jun Ki, Toscano, Robert Lopez, Stiehl, Walter D. and Breazeal, Cynthia (2008): The Design of a Semi-Autonomous Robot Avatar for Family Communication and Education. In: Buss, Martin and Kühnlenz, Kolja (eds.) Proceeding of 17th IEEE International Symposium on Robot and Human Interactive Communication RO-MAN August 1-3, 2008, Munich, Germany. pp. 166-173

Lohan, Katrin S., Pitsch, Karola, Rohlfing, Katharina J., Fischer, Kerstin, Saunders, Joe, Lehmann, H., Nehaniv, Christopher L. and Wrede, Britta (2011): Contingency allows the robot to spot the tutor and to learn from interaction. In: IEEE International Conference on Development and Learning ICDL 2011 August 24-27, 2011, Frankfurt, Germany. pp. 1-8

Lohse, Manja, Hanheide, Marc, Wrede, Britta, Walters, Michael L., Koay, Kheng L., Syrdal, Dag S., Green, Anders, Hüttenrauch, Helge, Dautenhahn, Kerstin, Sagerer, Gerhard and Severinson-Eklundh, Kerstin (2008): Evaluating extrovert and introvert behaviour of a domestic robot - a video study. In: Proceedings of the IEEE 17th International Symposium on Robot and Human Interactive Communication RO-MAN 2008 August 1-3, 2008, Munich, Germany. pp. 488-493

Lungarella, Max, Metta, Giorgio, Pfeifer, Rolf and Sandini, Giulio (2003): Developmental robotics: A Survey. In Developmental robotics: a survey, 15 (4) pp. 151-190

Lyon, Caroline, Nehaniv, Chrystopher L. and Saunders, Joe (2012): Interactive Language

*Learning by Robots: The Transition from Babbling to Word Forms. In PLoS One, 7 (6)*

MacDorman, Karl F. and Ishiguro, Hiroshi (2006): *The uncanny advantage of using androids in cognitive and social science research. In Interaction Studies, 7 (3) pp. 297-337*

Marti, Patrizia and Bannon, Liam J. (2009): *Exploring User-Centred Design in practice: Some caveats. In Knowledge, Technology & Policy, 22 (1) pp. 7-15*

Maulsby, David, Greenberg, Saul and Mander, Richard (1993): *Prototyping an Intelligent Agent through Wizard of Oz. In: Ashlund, Stacey, Mullet, Kevin, Henderson, Austin, Hollnagel, Erik and White, Ted (eds.) Proceedings of the ACM CHI 93 Human Factors in Computing Systems Conference April 24-29, 1993, Amsterdam, The Netherlands. pp. 277-284*

Melson, Gail F., Jr., Peter H. Kahn, Beck, Alan, Friedman, Batya, Roberts, Trace, Garrett, Erik and Gill, Brian T. (2009): *Children's behavior toward and understanding of robotic and living dogs. In Journal of Applied Developmental Psychology, 30 (2) pp. 92-102*

Metta, Giorgio, Natale, Lorenzo, Nori, Francesco, Sandini, Giulio, Vernon, David, Fadiga, Luciano, Hofsten, Claes von, Rosander, Kerstin, Lopes, Manuel, Santos-Victor, José, Bernardino, Alexandre and Montesano, Luis (2010): *The iCub Humanoid Robot: An Open-Systems Platform for Research in Cognitive Development. In Neural Networks, 23 (8) pp. 1125-1134*

Mueller, Florian, Vetere, Frank, Gibbs, Martin R., Kjeldskov, Jesper, Pedell, Sonja and Howard, Steve (2005): *Hug over a distance. In: Proceedings of ACM CHI 2005 Conference on Human Factors in Computing Systems 2005. pp. 1673-1676*

Mutlu, Bilge, Kanda, Takayuki, Forlizzi, Jodi, Hodgins, Jessica and Ishiguro, Hiroshi (2012): *Conversational gaze mechanisms for humanlike robots. In ACM Transactions on Interactive Intelligent Systems, 1 (2) p. 33*

Nehaniv, Chrystopher L. and Dautenhahn, Kerstin (eds.) (2007): *Imitation and Social Learning in Robots, Humans and Animals: Behavioural, Social and Communicative Dimensions. Cambridge, United Kingdom, Cambridge University Press*

Nehaniv, Chrystopher L., Förster, Frank, Saunders, Joe, Broz, Frank, Antonova, Elena, Köse, Hatice, Lyon, Caroline, Lehmann, Hagen, Sato, Yo and Dautenhahn, Kerstin (2013): *Interaction and Experience in Enactive Intelligence and Humanoid Robotics. In: In proceeding of IEEE Symposium on Artificial Life IEEE ALIFE, IEEE Symposium Series on Computational Intelligence IEEE SSCI 2013 April 15-19, 2013, Singapore,Singapore.*

Newell, Alan F., Carmichael, A., Morgan, M. and Dickinson, A. (2006): *The use of theatre in re-*

*quirements gathering and usability studies. In Interacting with Computers, 18 (5) pp. 996-1011*

*Nikolaidis, Stefanos and Shah, Julie (2013): Human-Robot Cross-Training: Computational Formulation, Modeling and Evaluation of a Human Team Training Strategy. In: Proceedings of the 8th ACM/IEEE international conference on Human-robot interaction March 3-6, 2013, Tokyo, Japan. pp. 33-40*

*Nolfi, Stefano and Floreano, Dario (2000): Evolutionary robotics: The biology, intelligence and technology of self-organizing machines. Cambridge, USA, MIT Press*

*O'Hare, Gregory M.P., Duffy, Brian R., Bradley, John F. and Martin, Alan N. (2003): Agent Chameleons: Moving Minds from Robots to Digital Information Spaces. In: Proceedings of Autonomous Minirobots for Research and Edutainment 2003. pp. 18-21*

*Papadopoulos, Fotios (2012). Socially Interactive Robots as Mediators in Human-Human Remote Communication. University of Hertfordshire http://uhra.herts.ac.uk/bitstream/handle/2299/9151/08171963%20Papadopoulos%20Fotios%20-%20final%20PhD%20submission.pdf;jsessionid=2854AA98DFF6A825C09F0D144F5ADBD2?sequence=1*

*Papadopoulos, Fotios, Dautenhahn, Kerstin and Ho, Wan Ching (2012): Exploring the use of robots as social mediators in a remote human-human collaborative communication experiment. In Paladyn- Journal of Behavioural Robotics, 3 (1) pp. 1-10*

*Parlitz, Christopher, Hägele, Martin, Klein, Peter, Seifert, Jan and Dautenhahn, Kerstin (2008): Care-o-Bot 3 - Rationale for Human-Robot Interaction Design. In: Proceedings of 39th International Symposium on Robotics ISR 2008 October 15-17, 2008, Seoul, South Korea. pp. 275-280*

*Pervez, Aslam and Ryu, Jeha (2008): Safe Physical Human Robot Interaction-Past, Present and Future. In Journal of Mechanical Science and Technology, 22 (3) pp. 469-483*

*Pfeifer, Rolf (2007): How the body shapes the way we think: a new view of intelligence. Cambridge, USA, MIT Press*

*Reeves, Byron and Nass, Clifford (1996): The media equation: How people treat computers, television and new media like real people and places. Cambridge University Press*

*Reiser, Ulrich, Parlitz, Christopher and Klein, Peter (2013b): Care-O-bot® 3 – Vision of a robot butler. In: Trappl, Robert (ed.). "Your Virtual Butler - The Making-of". Springer Linkpp. 97-116*

*Reiser, Ulrich, Jacobs, Theo, Arbeiter, Georg, Parlitz, Christopher and Dautenhahn, Kerstin (2013a): Care-O-bot® 3 - Vision of a Robot Butler. Your Virtual Butler - The Making-of. In: Trappl, Robert (ed.). "Lecture Notes in Artificial Intelligence". Berlin, Germany: Springer*

*Linkpp. 97-116*

Robins, Ben, Dautenhahn, Kerstin and Dubowski, Janek (2004): *Investigating Autistic Children's Attitudes Towards Strangers with the Theatrical Robot - A New Experimental Paradigm in Human-Robot Interaction Studies. In: 13th IEEE International Workshop on Robot and Human Interactive Communication, ROMAN 2004 September 20-22, 2004, Kurashiki, Okayama Japan. pp. 557-562*

Robins, Ben, Dautenhahn, Kerstin and Dubowski, Janek (2005): *Robots as isolators or mediators for children with autism A cautionary tale. In: Proceedings of the Symposium on Robot Companions Hard Problems and Open Challenges in Human-Robot Interaction AISB 05 April 14-15, 2005, Hatfield, United Kingdom. pp. 82-88*

Robins, Ben, Ferrari, Ester, Dautenhahn, Kerstin, Kronreif, Gernot, Prazak-Aram, Barbara, Gelderblom, Gert-jan, Tanja, Bernd, Caprino, Francesca, Laudanna, Elena and Marti, Patrizia (2010): *Human-centred design methods: Developing scenarios for robot assisted play informed by user panels and field trials. In International Journal of Human-Computer Studies, 68 (12) pp. 873-898*

Robins, B., Dautenhahn, K., Boekhorst, R. Te and Billard, A. (2005a): *Robotic assistants in therapy and education of children with autism: can a small humanoid robot help encourage social interaction skills?. In Universal Access in the Information Society, 4 (2) pp. 105-120*

Robins, Ben, Dautenhahn, Kerstin and Dickerson, Paul (2009): *From Isolation to Communication: A Case Study Evaluation of Robot Assisted Play for Children with Autism with a Minimally Expressive Humanoid Robot. In: Proceedings of the 2009 International Conference on Advances in Computer-Human Interactions 2009. pp. 205-211*

Robins, Ben, Dautenhahn, Kerstin, Ferrari, Ester, Kronreif, Gernot, Prazak-Aram, Barbara, Marti, Patrizia, Iacono, Iolanda, Gelderblom, Gert Jan, Bernd, Tanja, Caprino, Francesca and Laudanna, Elena (2012): *Scenarios of robot-assisted play for children with cognitive and physical disabilities. In Interaction Studies, 13 (2) pp. 189-234*

Rossano, Federico (2013): *Sequence organization and timing of bonobo mother-infant interactions. In Interaction Studies, 14 (2) pp. 160-189*

Sabanovic, Selma, Michalowski, Marek P. and Simmons, Reid (2006): *Robots in the Wild: Observing human-robot social interaction outside the lab. In: Proceedings of the 9th International Workshop on Advanced Motion Control AMC 2006 March 27-29, 2006, Istanbul, Turkey. pp. 576-581*

Salem, Maha, Kopp, Stefan, Wachsmuth, Ipke, Rohlfing, Katharina and Joublin, Frank

(2012): Generation and Evaluation of Communicative Robot Gesture. In International Journal of Social robotics, 4 (2) pp. 201-217

Salter, Tamie, Michaud, François and Larouche, Hélène (2010): How wild is wild A taxonomy to categorize the wildness of child-robot interaction. In International Journal of Social robotics, 2 (4) pp. 405-415

Santis, Agostino De, Siciliano, Bruno, Luca, Alessandro De and Bicchi, Antonio (2008): An atlas of physical human-robot interaction. In Mechanism and Machine Theory, 43 (3) pp. 253-270

Scassellati, Brian, Admoni, Henny Y. and Matarić, Maja J. (2012): Robots for Use in Autism Research. In Annual Review of Biomedical Engineering, 14 pp. 275-294

Scholtz, Jean (2003): Theory and Evaluation of Human Robot Interactions. In: Proceedings of the 36th Annual Hawaii International Conference on System Sciences HICSS03 January 6-9, 2003, Hawaii, USA. p. 125

Schulz, R., Beach, S.R., Matthews, J. Tabolt, Courtney, K.L. and Darbbs, A.J. De Vito (2012): Designing and evaluating quality of life technologies: An interdisciplinary approach. In Proceedings of the IEEE, 100 (8) pp. 2397-2409

Sharkey, Amanda and Sharkey, Noel (2011): Children, the Elderly, and Interactive Robots. In IEEE Robotics and Automation Magazine, 18 (1) pp. 32-38

Sharkey, Amanda and Sharkey, Noel (2012): Granny and the robots: ethical issues in robot care for the elderly. In Ethics and Information Technology, 14 (1) pp. 27-40

Shen, Qiming, Kose-Bagci, Hatice, Saunders, Joe and Dautenhahn, Kerstin (2011): The Impact of Participants' Beliefs on Motor Interference and Motor Coordination in Human-Humanoid Interaction. In IEEE Trends in Autonomous Mental Development, 3 (1) pp. 6-16

Shibata, Takanori, Kawaguchi, Yukitaka and Wada, Kazuyoshi (2012): Investigation on People Living with Seal Robot at Home - Analysis of Owners' Gender Differences and Pet Ownership Experience. In International Journal of Social robotics, 4 (1) pp. 56-63

Shimada, Michihiro and Kanda, Takayuki (2012): What is the appropriate speech rate for a communication robot. In Interaction Studies, 13 (3) pp. 408-435

Sung, Ja-Young, Guo, Lan, Grinter, Rebecca E. and Christensen, Henrik I. (2007): "My Roomba Is Rambo": Intimate Home Appliances. In: Krumm, John, Abowd, Gregory D., Seneviratne, Aruna and Strang, Thomas (eds.) UbiComp 2007 Ubiquitous Computing - 9th International Conference September 16-19, 2007, Innsbruck, Austria. pp. 145-162

Sung, Ja-Young, Grinter, Rebecca E., Christensen, Henrik I. and Guo, Lan (2008): Housewives or technophiles?: understanding domestic robot owners. In: Proceedings of the 3rd ACM/IEEE International Conference on Human Robot Interaction 2008. pp. 129-136

Syrdal, Dag Sverre, Otero, Nuno and Dautenhahn, Kerstin (2008): Video Prototyping in Human-Robot Interaction: Results from a Qualitative Study. In: Abascal, Julio, Fajardo, Inmaculada and Oakley, Ian (eds.) Proceedings of the European Conference on Cognitive Ergonomics 2008 September 16-19, 2008, Madeira, Portugal. pp. 132-140

Syrdal, Dag Sverre, Dautenhahn, Kerstin, Walters, Michael L., Koay, Kheng Lee and Otero, Nuno R. (2011): The Theatre methodology for facilitating discussion in human-robot interaction on information disclosure in a home environment. In: Proceedings RO-MAN 2011, 20th IEEE International Symposium on Robot and Human Interactive Communication July 31- August 3, 2011, Georgia, USA. pp. 479-484

Syrdal, Dag Sverre, Koay, Kheng L., Gácsi, Márta, Walters, Michael L. and Dautenhahn, Kerstin (2010): Video Prototyping of Dog-Inspired Non-verbal Affective Communication for an Appearance Constrained Robot. In: Proceedings IEEE RO-MAN 2010, 19th IEEE International Symposium in Robot and Human Interactive Communication September 12-15, 2010, Viareggio, Italy. pp. 632-637

Teh, James Keng Soon, Cheok, Adrian David, Peiris, Roshan L., Choi, Yongsoon, Thuong, Vuong and Lai, Sha (2008): Huggy Pajama: a mobile parent and child hugging communication system. In: Proceedings of ACM IDC08 Interaction Design and Children 2008. pp. 250-257

Thomaz, Andrea L. and Breazea, Cynthia (2008): Teachable robots: Understanding human teaching behavior to build more effective robot learners. In Artificial Intelligence, 172 (6) pp. 716-737

Thomaz, Andrea L. and Cakmak, Maya (2009): Learning about objects with human teachers. In: Proceedings of the 4th ACM/IEEE International Conference on Human Robot Interaction 2009. pp. 15-22

Tolman, Edward C. (1948): Cognitive Maps in rats and men. In The Psychological Review, 55 (4) pp. 189-208

Turkle, Sherry (2007): Authenticity in the age of digital companions. In Interaction Studies, 8 (3) pp. 501-517

Turkle, Sherry, Taggart, Will, Kidd, Cory D. and Dasté, Olivia (2006): Relational artifacts with children and elders: the complexitites of cybercompanionship. In Connection Science, 18 (4) pp. 347-361

Turner, Mark (1996): The Literary Mind: The Origins of Thought and Language. New York, USA, Oxford University Press

Vernon, David, Hofsten, Claes von and Fadiga, Luciano (2011): A Roadmap for Cognitive Development in Humanoid Robots. Berlin,Germany, Springer

Wada, Kazuyoshi and Shibata, Takanori (2007): Living with seal robots in a care house-Evaluations of social and physiological influences. In: IEEE/RSJ International Conference on Intelligent Robots and Systems, 2006 October 9-15, 2007, Beijing, China. pp. 4940-4945

Walters, Michael L., Lohse, Manja, Hanheide, Marc, Wrede, Britta, Syrdal, Dag Sverre, Koay, Kheng L., Green, Anders, Hüttenrauch, Helge, Dautenhahn, Kerstin, Sagerer, Gerhard and Eklundh, Kerstin Severinson (2011): Evaluating the behaviour of domestic robots using video-based studies. In Advanced Robotics, 25 (18) pp. 2233-2254

Walters, Michael L., Syrdal, Dag S., Dautenhahn, Kerstin, Boekhorst, Rene te and Koay, Kheng L. (2008): Avoiding the uncanny valley: robot appearance, personality and consistency of behavior in an attention-seeking home scenario for a robot companion. In Autonomous Robots, 24 (2) pp. 159-178

Webb, Barbara (2001): Can robots make good models of biological behaviour?. In Behavioural and Brain Sciences, 24 (6) pp. 1033-1050

Weir, Sylvia and Emanuel, Ricky (1976). Using LOGO to catalyse communication in an autistic child. University of Edinburgh

Werry, Iain and Dautenhahn, Kerstin (2007): Human-Robot Interaction as a Model for Autism Therapy: An Experimental Study with Children with Autism. In: Laubichler, Manfred D. and Müller, Gerd B. (eds.). "Modeling Biology: Structures, Behaviors, Evolution (Vienna Series in Theoretical Biology". Massachusetts, USA: MIT Presspp. 283-299

Werry, Iain, Dautenhahn, Kerstin, Ogden, Bernard and Harwin, William (2001): Can Social Interaction Skills Be Taught by a Social Agent The Role of a Robotic Mediator in Autism Therapy. In: Nehaniv, Chrystopher L., Dautenhahn, Kerstin and Beynon, Meurig (eds.) Proceedings of the 4th International Conference on Cognitive Technology Instruments of Mind CT2001 August 6-9, 2001, Warwick, United Kingdom. pp. 57-74

Wing, Lorna (1996): The Autistic Spectrum. Constable Press

Woods, Sarah N., Walters, Michael L., Koay, Kheng Lee and Dautenhahn, Kerstin (2006): Methodological Issues in HRI: A Comparison of Live and Video-Based Methods in Robot to Human Approach Direction Trials. In: Proceedings of The 15th IEEE International

*Symposium on Robot and Human Interactive RO-MAN06 September 6-8, 2006, Hatfield, United Kingdom. pp. 51-58*

*Woods, Sarah, Dautenhahn, Kerstin, Kaouri, Christina, Boekhorst, Rene te, Koay, Kheng Lee and Walters, Michael L. (2007): Are Robots Like People - Relationships between Participant and Robot Personality Traits in Human-Robot Interaction Studies. In Interaction Studies, 8 (2) pp. 281-305*

*Woods, Sarah N., Walters, Michael L., Koay, Kheng Lee and Dautenhahn, Kerstin (2006): Comparing Human Robot Interaction Scenarios Using Live and Video Based Methods: Towards a Novel Methodological Approach. In: The 9th IEEE International Workshop on Advanced Motion Control AMC06 March 27-29, 2006, Istanbul, Turkey. pp. 770-775*

*Yamaoka, Fumitaka, Kanda, Takayuki, Ishiguro, Hiroshi and Hagita, Norihiro (2007): How contingent should a lifelike robot be? The relationship between contingency and complexity. In Connection Science, 19 (2) pp. 143-162*

*Yamazaki, Kimitoshi, Ueda, Ryohei, Nozawa, Shunichi, Kojima, Mitsuharu, Okada, Kei, Matsumoto, Kiyoshi, Ishikawa, Masaru, Shimoyama, Isao and Inaba, Masayuki (2012): Home-assistant robot for an aging society. In Proceedings of the IEEE, 100 (8) pp. 2429-2441*

## About the author

### Kerstin Dautenhahn



I am Professor of Artificial Intelligence in the School of Computer Science, Faculty of Engineering and Information Sciences at University of Hertfordshire, where I am a coordinator of the Adaptive Systems Research Group. Main areas of my research are Human-Robot Interaction, Social Robotics, Socially Intelligent Agents and Artificial Life, you can find out more about it on this homepage. I am former member of the Department of Biological Cybernetics at University of Bielefeld, Germany, 1990-1993, and AI-Lab at GMD, Sankt Augustin, Germany, 1993-1996, and VUB Artificial Intelligence Laboratory, Brussels, Belgium, until end of 1996. From January 1997 to April 2000 Lecturer, Department of Cybernetics at University of Reading, United Kingdom. In April 2000 I joined the Department of Computer Science (now School of Computer Science) at University of Hertfordshire as Principal Lecturer. Later I got promoted to Reader and then Research Professor. I am Editor in Chief (jointly with Prof. Angelo Cangelosi - University of Plymouth) of the Journal Interaction Studies- Social Behaviour and Communication in Biological and Artificial Systems published by John Benjamins Publishing Company, Associate Editor of Adaptive Behavior, Sage Publications, Associate Editor of the International Journal of Social Robotics, published by Springer and Associate Editor of the IEEE Transactions on Autonomous Mental Development. I am an Editor of the book series Advances in Interaction Studies, published by John Benjamins Publishing Company. I am a Fellow of the Royal Society for the encouragement of Arts, Manufactures and Commerce (RSA), as well as a member of IEEE, ACM, SSAISB, and the German organisations GI and GK.

## Your notes and thoughts on chapter 38

Record your notes and thoughts on this chapter. If you want to share these thoughts with others online, go to the bottom of the page at:
http://www.interaction-design.org/encyclopedia/human-robot_interaction.html

**NOTES:**

# CHAPTER 40

# Emotion and Website Design

***by Dianne Cyr.***

This chapter is about hedonic or affective elements[1] of website design and the potential of such design to elicit emotion in the user. In an online environment hedonic elements of website design include color, images, shapes, and use of photographs, among other characteristics, which are expected to provide the user with emotional appeal, a sense of the aesthetic, or a positive impression resulting from the overall graphical look of a website (Cyr et al., 2009; Lavie and Tractinsky, 2003; Zhang, 2013). While it is well known that emotion is important to the interpretation of experience, it is only in recent years that research has begun to transcend utilitarian aspects of website design to consider empirically affective elements of design. Therefore, not only is it important that websites are useful and easy to use, but also that they entice the user to experience emotions such as enjoyment, involvement, trust, or satisfaction.

1. In this article, the terms affective and hedonic are used interchangeably.

In this context, I focus on an empirically based research perspective that is anchored in the tradition of information systems (IS), and more specifically on the design of websites in the human computer interaction tradition (HCI). While much of the research considered is anchored in e-commerce, there are clearly implications for other types or applications such as e-government or social networking. In the following pages I therefore include the following: a brief retrospective examination of the development of a hedonic perspective in IS and design; an outline of some of the more commonly documented emotion-laden outcomes of website design; a consideration of graphical design elements known to elicit emotion in the user such as human images and color; an elaboration on the social elements of design; and a conclusion with a segment on future directions for research.

## 40.1 EMOTION AND WEBSITE DESIGN: SOME BACKGROUND

### 40.1.1 Defining Emotion

According to Zhang, 2013 ( p. 247) "[A]ffect is conceived of as an umbrella term for a set of more specific concepts that includes emotion, moods, feelings..." Zhang continues (ibid, p. 251):

> *"One of the most complex affective concepts is emotion. Put simply, emotions are induced affective states (Clore and Schnall 2005), or core affect attributed to stimuli (Barrett et al. 2007; Russell 2003). Emotions typically arise as reactions to situational events in an individual's environment that are appraised to be relevant to his/her needs, goals, or concerns. Once activated, emotions generate subjective feelings (such as anger or joy), generate motivational states with action tendencies, arouse the body with energy-mobilizing responses that prepare it for*

> *adapting to whatever situation one faces, and express the quality and intensity of emotionality outwardly and socially to others*
>
> *(Damasio 2001; Izard 1993; Reeve 2005)"*

Other researchers have identified a spectrum of feeling associated with emotion. For example, different emotions included by different authors have been anger, guilt, sadness, and fear/anxiety (Smith and Lazarus, 1993), or joy, fear, anger, sadness, disgust, shame, and guilt (Scherer, 1997). Emotion, therefore, has been seen to have both negative and positive valence (Cenfetelli, 2004; Roseman et al., 1996). Emotional responses are known to have two components: arousal and valence. Arousal reflects the intensity of the response, while valence refers to the direct emotional response ranging from positive to negative (Russell, 1980; Deng and Poole, 2010).

With website design, it is expected that emotion is aroused in the user based on a response to specific design elements. Therefore, the user may feel a sense of satisfaction when website colors are appealing, or when a graphical design elicits enjoyment or excitement. In addition, it is important that website design meets the needs and sensibilities of the user. This may include website design that is particular to subgroups with specific preferences. The author's research has identified that there are different website design preferences for men and women, or for users in different national locations. These differences will be elaborated in the following pages. If website design is appropriate to the user, then this arouses the action tendencies described by (Zhang (2013); above, including users being more loyal to the site, and returning there in the future.

### 40.1.2 Beyond Cognitive-based Paradigms

Despite the pervasiveness of emotional reaction in the human psyche, only within the last decade have calls been made for a break with conventional cognition-driven paradigms of studying user reactions to technology (Beaudry and Pinsonneault, 2010; Zhang and Li, 2004). In their place is an expanded focus that includes not only utilitarian outcomes such as usefulness or ease of use, but also the role of affect and emotion in the examination of information and communication technology systems (Kim et al., 2007; Sun and Zhang, 2006). For instance, Hassenzahl 2006 (p. 266) elaborates:

> *"In HCI, it is widely accepted that usability is the appropriate definition of quality. However, the focus of usability on work-related issues (e.g., effectiveness, efficiency) and cognitive information processing has been criticized. Its quite narrow definition of quality neglects additional hedonic (non-instrumental) human needs and related phenomena, such as emotion, affect and experience"*

Further, Zhang (2013) outlines:

> *Affect is a critical factor in human decisions and behaviors within many social contexts. In the information and communication technology context (ICT), a growing number of studies consider the affective dimension of human interaction with ICTs. However, few of these studies take systematic approaches, resulting in inconsistent conclusions and contradictory advice for researchers and practitioners.*

To date, some research has been conducted in the area of emotion when users are in online environments. For instance, hedonic outcomes have been examined in terms of flow (Eroglu et al., 2003; Griffith et al., 2001; Ha et al., 2007; Huang, 2006; Koufaris et al., 2002); cognitive absorption (Agarwal and Karahanna, 2000; Wakefield and Whitten, 2006), involvement (Fortin and Dholakia, 2003; Johnson et al., 2006), playfulness (Wakefield and Whitten, 2006), enjoyment (Dickinger et al., 2008; Lee et al., 2000; Li et al., 2008; Lin and Bhattarcherjee, 2008;2010; Sun, 2001; Sun and Zhang, 2006; van der Heijden, 2004; Venkatesh, 2000; Wakefield and Whitten, 2006), hedonic outcomes (Venkatesh and Brown, 2001), pleasure (Belanger et al., 2002), happiness (Beadry and Pinsonneault 2010), fun (Dabholkar, 1994; Dabholkar and Bagozzi, 2002), stimulation (Fiore et al., 2005), or mystery (Rosen and Purinton, 2004), among others.

Further, various studies are emerging that examine emotion more specifically related to design elements, including design of e-commerce websites. This is important since Lam and Lin (2004) argue that the role of emotions in online shopping is even more important than in traditional marketing contexts because the consumer is disengaged from human interaction. To this end, user emotional responses have been measured with respect to “design factors” such as shapes, textures, color (Kim et al., 2003), visual characteristics of web pages (Lindgaard et al., 2006), or web page aesthetics (Robins and Holmes, 2008). Additional topics covered are affective user interfaces (Johnson and Wiles, 2003; Lisetti and Nasoz, 2002); hedonic quality (Childers et al., 2001; Hassenzahl, 2002; van der Heijden 2003, 2004); aesthetic performance including atmospheric cues, media richness and social presence (Lim and Cyr, 2009); presentation richness such as symbol variety (Jahng et al., 2002); interaction richness (Jahng et al., 2007); human images (Cyr et al., 2009); color (Cyr et al., 2010); or vividness (Jiang and Benbasat, 2007). Cyr et al. (2006) found that design aesthetics on a mobile device resulted in enjoyment, and ultimately online loyalty. Similarly, Sony Ericsson’s website for Egypt (Figure 40.1) is aimed to promote variety and fun (Seidenspinner and Theuner, 2007).



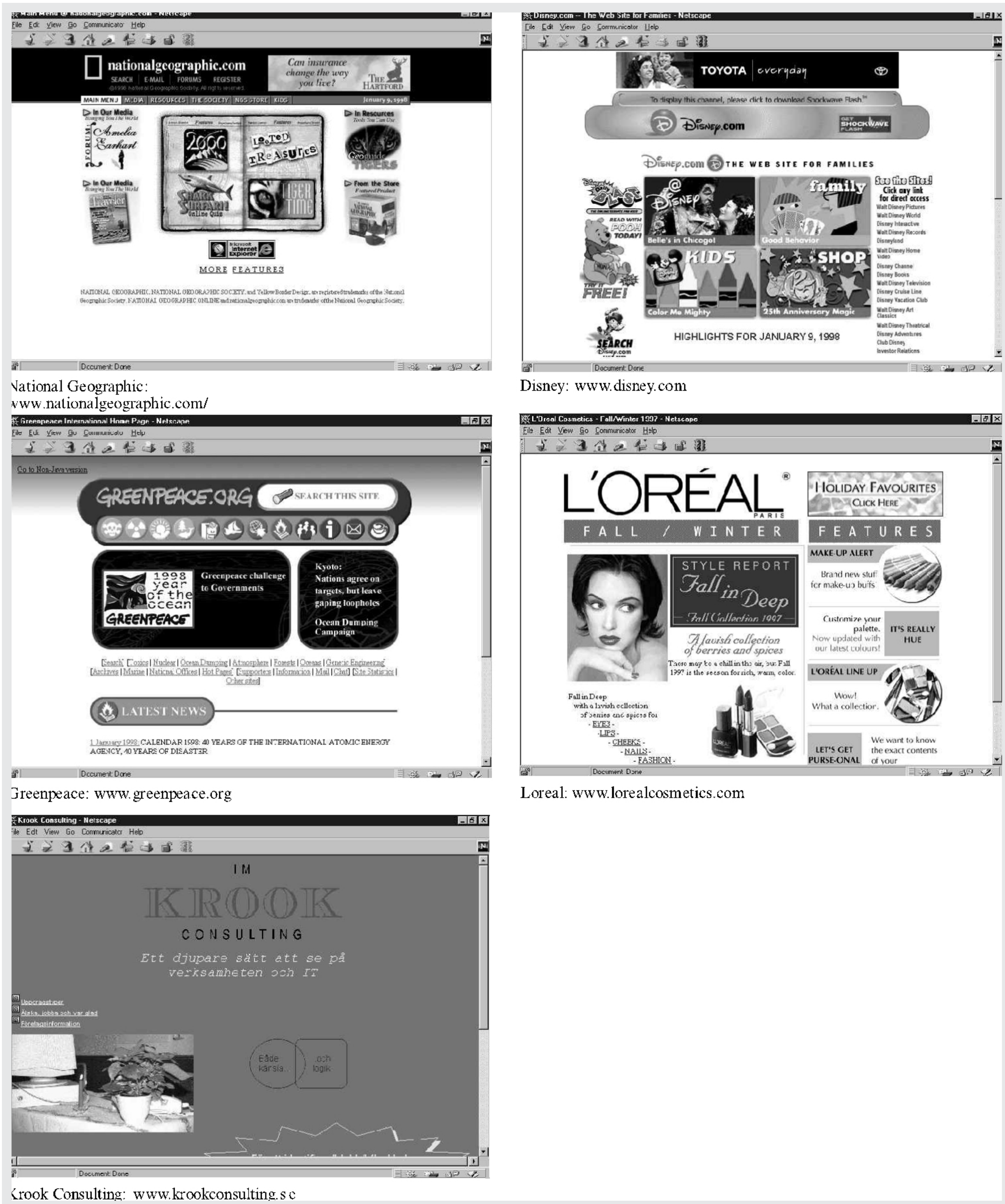

FIGURE 40.1: Sony Ericsson Egypt: Promoting Variety and Fun (in Seidenspinner and Theuner, 2007).

A note on the use of illustrations in this chapter: In many cases these figures are reproduced from the original article where they appeared. These figures are used for general illustrative purposes only, and it is not intended that full readability is required.

Over the years, beauty has been a precursor to emotional responses. In website design various researchers have focused on aesthetic beauty[2] (e.g. Karvonen,

2. A thorough discussion of aesthetics is beyond the scope of this article, although it may be of interest to the reader. Refer to Lavie and Tractinsky (2004) for a discussion of "classical aesthetics" and "expressive aesthetics", and also to the overview article by Tractinsky (2012) titled Visual Aesthetics in Visual Aesthetics in Human Computer Interaction and Design, In Soegaard, M. and Dam, R.F. (Eds.) Encyclopedia of Human-Computer Interaction and Interaction Design. Aarhus, Denmark: The Interaction Design Foundation. Available online at:http://www.interaction-design.org/encyclopedia/visual_aesthetics.html

2000; Lavie and Tractinsky, 2004). Schenkman and Jonsson (2000)examined beauty on a sample of web pages (Figure 40.2) using multi-dimensional analysis. They found four categories that viewers used to judge a web page: beauty; illustrations versus text; overview (i.e. lucid, clear, and easy to understand); and structure. Overall, the researchers found the best predictor of overall judgment of the website was beauty. Based on user judgments, these websites were scored for perceived beauty out of a maximum of seven: National Geographic (5.08), Disney (4.51), Greenpeace (3.93), L'Oreal (5.55), and Krook Consulting (2.70).




*Courtesy of National Geographic, Disney, Greenpeace, L'Oreal Cosmetics, and Krook Consulting. Copyright: compositeWorkWithMultipleCopyrightTerms (Work that is derived from or composed of multiple works with varying copyright terms and/or copyright holders).*

FIGURE 40.2: Sample Web Pages related to Beauty.(in Schenkman and Jonsson, 2000).

Deng and Poole (2010) examined web page visual complexity and order and found a relationship to user emotions and behavior. Website complexity was dependent on the number of links, graphics, and the amount of text that appear on the web page. Figure 40.3 shows examples of low website complexity (12 links/2 graphics/33 text) versus high website complexity shown in Figure 40.4 (54 links/14 graphics/118 text).

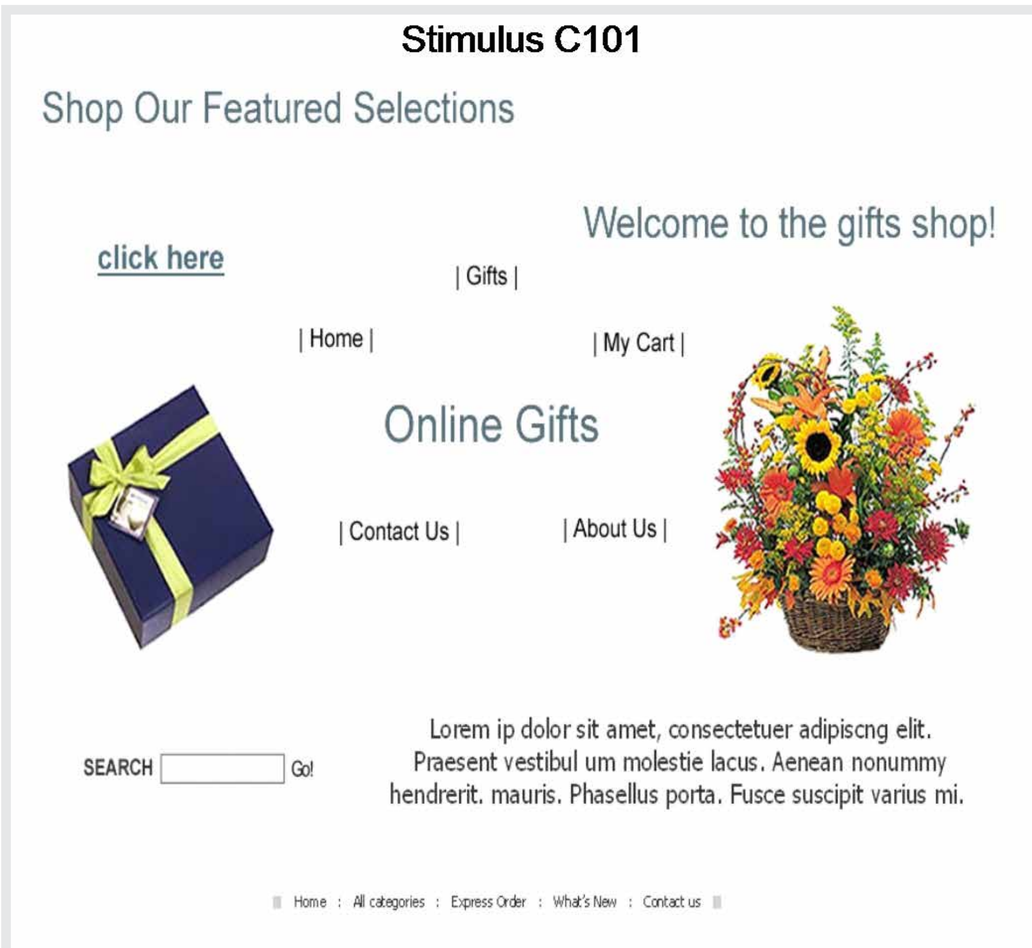





Figure 40.3: Sample of a Web Page Low in Visual Complexity (in Deng and Poole, 2010).




FIGURE 40.4: Sample of a Web Pages High in Visual Complexity (in Deng and Poole, 2010).

Order refers to the logical organization, coherence, and clarity of the web page content. An additional component of the research is the degree to which the meta-motivational state of the user (e.g. extent to which a user seeks stimulation from the site) influences the user's impression of the website. It was found that whether a user was more focused or more relaxed in approach when viewing a website did have an effect on whether the website was perceived as pleasant. This second finding suggests that not only is the actual design of the website important to eliciting an emotional reaction from the user, but also that the user's motivational state will have some bearing on how the website is viewed and evaluated.

Scholars have also examined how social elements such as pictures of people or emotive text on websites empirically impact users' impressions such as enjoyment (Cyr et al. 2006; Gefen and Staub 2003; Hassanein and Head, 2007).[3] Many of these studies, particularly in the e-commerce realm, have user outcomes of trust (Bhattacherjee, 2002; Chen and Dhillon, 2003; Cheung and Lee, 2006; Cyr, 2008; Everard and Galletta, 2006; Gefen et al., 2003; Jarvenpaa et al., 2000;Komiak and Benbasat, 2004 ; Koufaris and Hampton-Sosa, 2004; Wang and Benbasat, 2005) and satisfaction (Agarwal and Venkatesh, 2002; Fogg et al., 2002; Hoffman and Novak, 1996; Koufaris, 2002; Lindgaard and Dudek, 2003; Nielsen, 2001; Palmer, 2002; Szymanski and Hise, 2000; Yoon, 2002).

Further research has aimed to develop models that incorporate hedonic elements. Related to affect, Loiacono and Djamasbi (2010) proposed the relevance of mood (such as sadness, fear, or happiness) for system usage models that could be applied online. They further outlined a model in which mood is intended to influence perception, evaluation, and cognitive effort resulting in variable levels of IS usage behavior. While this model is not tested, it is a useful framework from which to investigate emotion empirically.

In addition, Lowry et al. (forthcoming 2013) developed a hedonic system adoption model focused on a user's intrinsic motivations. Using an immersive

3. For a useful review of a cognitive-affective model of organizational communication that is broader that website design refer to Te'eni (2001).

gaming environment, various games were rigorously tested to determine perceived levels of joy and ease of use. A game that scored "low" on both dimensions was a text-based adventure game with minimal graphical content (Figure 40.5).

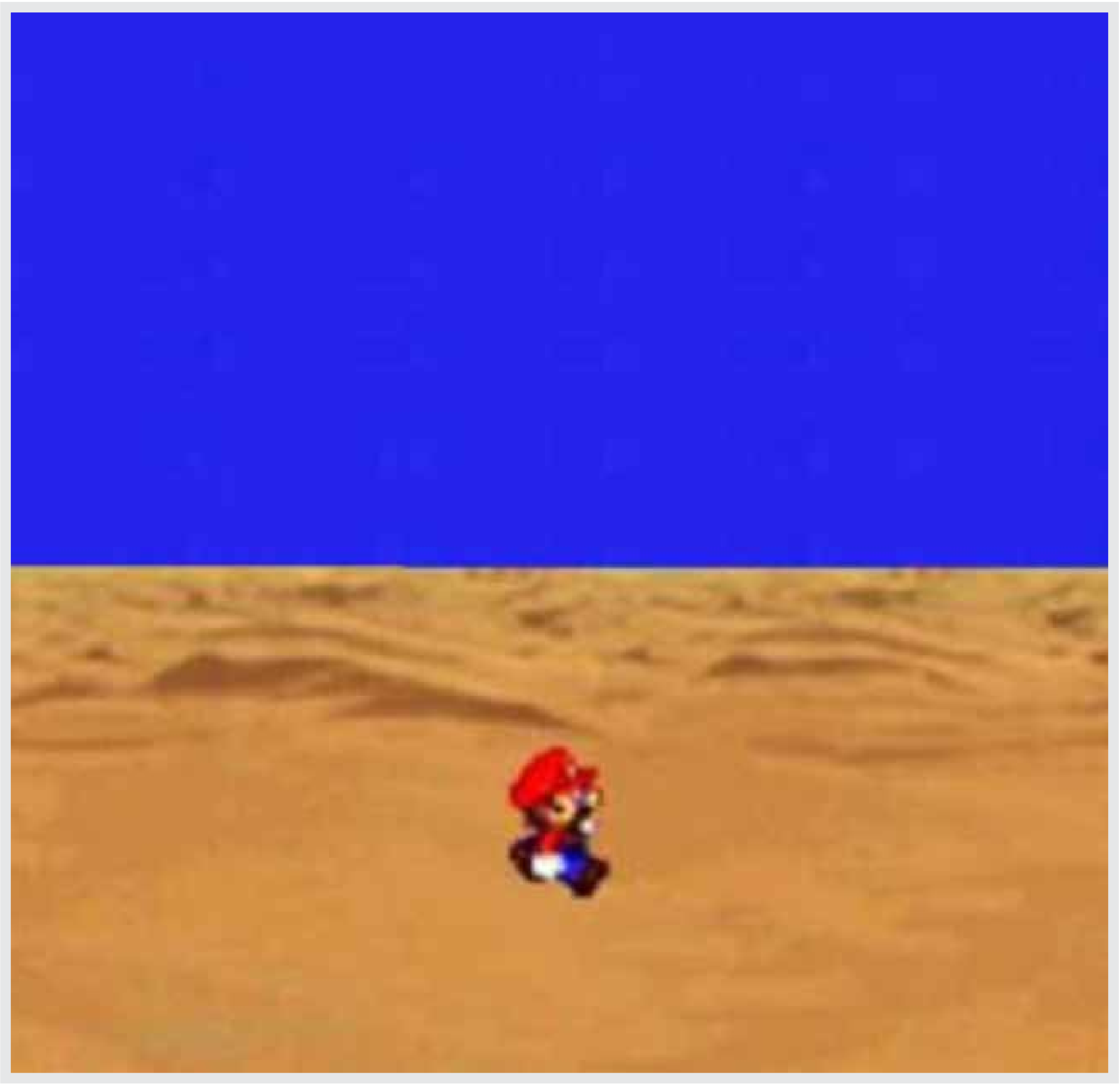




Figure 40.5: Gaming Environment Low in Hedonic Value (from Lowry et al., 2013).

In contrast, the game scored by users as "high" for joy and ease of use was highly interactive, with complex colors and graphics (Figure 40.6). Using various game interfaces, these researchers determined that perceived ease of use resulting in behavioral intention to use, or the user's experience of immersion, is mediated by hedonic constructs such as curiosity or joy (along with perceived usefulness and control). Hence the role of affectively based constructs is central to understanding the user experience in a variety of web-based platforms, including gaming.




Figure 40.6: Gaming Environment High in Hedonic Value (from Lowry et al., 2013).

### 40.1.3 Summary

It is encouraging that over the past decade, researchers have expanded beyond utilitarian models of online experience to encompass how users are also emotionally engaged. While substantial progress has been made in this area, only recently have researchers created comprehensive models to evaluate how hedonic systems operate. Hence, there is considerable scope for future work that empirically explores variables that elicit hedonic or affective responses in the user, with a goal of theory development in this domain. The spectrum for such research is broad, and encompasses e-commerce, gaming, e-health, and many additional contexts.

In general, emotional responses are triggered by an ability to engage the user in an online environment which is aesthetically pleasing. This places the elicitation of emotion firmly in the realms of visual design and interaction design. However, there is no clear definition as to what represents a hedonic outcome—and, based on the literature to date, these outcomes have varied widely. As such, there is scope for studies that add clarity and consistency to how hedonic systems impact the user. In an effort to consolidate the literature to date, in the following sections I have chosen to discuss four hedonic outcomes: enjoyment, involvement, trust, and satisfaction. These have particular importance in an e-commerce setting, but are relevant in online contexts more generally. These constructs have received considerable use and have been validated in numerous studies.

## 40.2 OUTCOME VARIABLES THAT ELICIT EMOTION

### 40.2.1 Enjoyment

As early as 2003, Blythe and Wright (2003) (p. xvi) argued that in HCI "traditional usability approaches are too limited and must be extended to encompass enjoyment". Perhaps more than any other construct, enjoyment has been used to measure user hedonic perceptions and expectations on websites (e.g. Dellaert and Dabholkar, 2009; Gretzel and Fersnemaier, 2006;Hassanein and Head, 2005 ; Koufaris et al., 2001; Koufaris, 2002; Lee et al., 2000; Li et al., 2008; Sun, 2001; Sun and Zhang, 2006;Qui and Benbasat, 2009; Venkatesh, 2000). Other work (e.g. Warner, 1980[4]) has suggested enjoyment encompasses three dimensions: engagement, positive affect, and fulfillment. Enjoyment has also been subsumed under the concept of flow (as originally identified by Csikszentmihalyi, 1989[5]). Although Dabholkar (1994) and Dabholkar and Bagozzi (2002) employed the term "play" in place of enjoyment in their research, they admitted that the meaning of "play" is no different from that of enjoyment.[6]

Although enjoyment is a commonly used construct to measure user reactions to hedonic content on the web, it is surprising that the accurate measurement of enjoyment has tended to be elusive. For example, as recently as 2008, Lin et al. 2008(p. 41) noted: "[W]hen we came to the question of assessing the degree to which enjoyment arises from a Web encounter, we found remarkably little to guide us...and no instrument for assessing enjoyment of Web experiences could be found." To this end, they created and validated an instrument for online enjoyment with three dimensions: engagement, positive affect, and fulfilment, as suggested earlier by Warner (1980). This instrument is a positive step forward to create ways in which to accurately measure user responses, and

4. For a thorough review of enjoyment in a non-Internet context refer to Warner (1980).

5. Csikszentmihalyi (1990) is the key figure in the development of the flow construct. Flow is comprised of four dimensions: control, focused attention, aroused curiosity, and intrinsically interested. In an online context, Hsu and Lu (2004) examined flow, which they characterize as an extremely enjoyable experience that includes total involvement, enjoyment, control, concentration, and intrinsic interest.

6. Although not specifically enjoyment, as early as 1994 Babin et al. developed a scale for the measurement of hedonic and utilitarian value when shopping (although not online). Of interest, three of the items in this scale of 11 items are directly related to enjoyment. Other single items relate to joy, escape, or excitement.

thus inform web managers and markets as to what constitutes meaningful and enjoyable website design.

Further, the concept of enjoyment has been revealed to be "a strong predictor of attitude in the web-shopping context" (Childers et al., 2001, p. 526; Cyr et al., 2006, 2007; Hassanein and Head, 2006; Lankton and Wilson, 2007; Koufaris et al., 2002; van der Heijden, 2003; Zhang and von Dran, 2002). In online settings, a primary goal of vendors is to entice users to purchase from websites or to revisit them in the future, resulting in loyal behavior (Rosen and Purinton, 2004). Online loyalty (or e-loyalty) has been described as an enduring psychological attachment by a customer to a particular online vendor or service provider (Anderson and Srinivasan, 2003; Butcher et al., 2001). Jiang and Benbasat (2007) discovered that vividness and interaction of consumer product displays for a watch and Personal Data Assistant (PDA) resulted in enjoyment and e-loyalty.

In a study of mobile interfaces as used in an e-service shopping environment, researchers found the "design aesthetics" of the interface positively impacted enjoyment, usefulness, and ease of use, which in turn positively affected user loyalty (Cyr et al., 2006). More specifically, Design Aesthetics referred to the following: attractiveness of the screen design (e.g. colors, boxes, menus); professional design; meaningful graphics; and overall "look and feel" as visually appealing (Figure 40.7).




Figure 40.7: Screen Shots used to Test Design Aesthetics (in Cyr et al., 2009).

### 40.2.2 Involvement

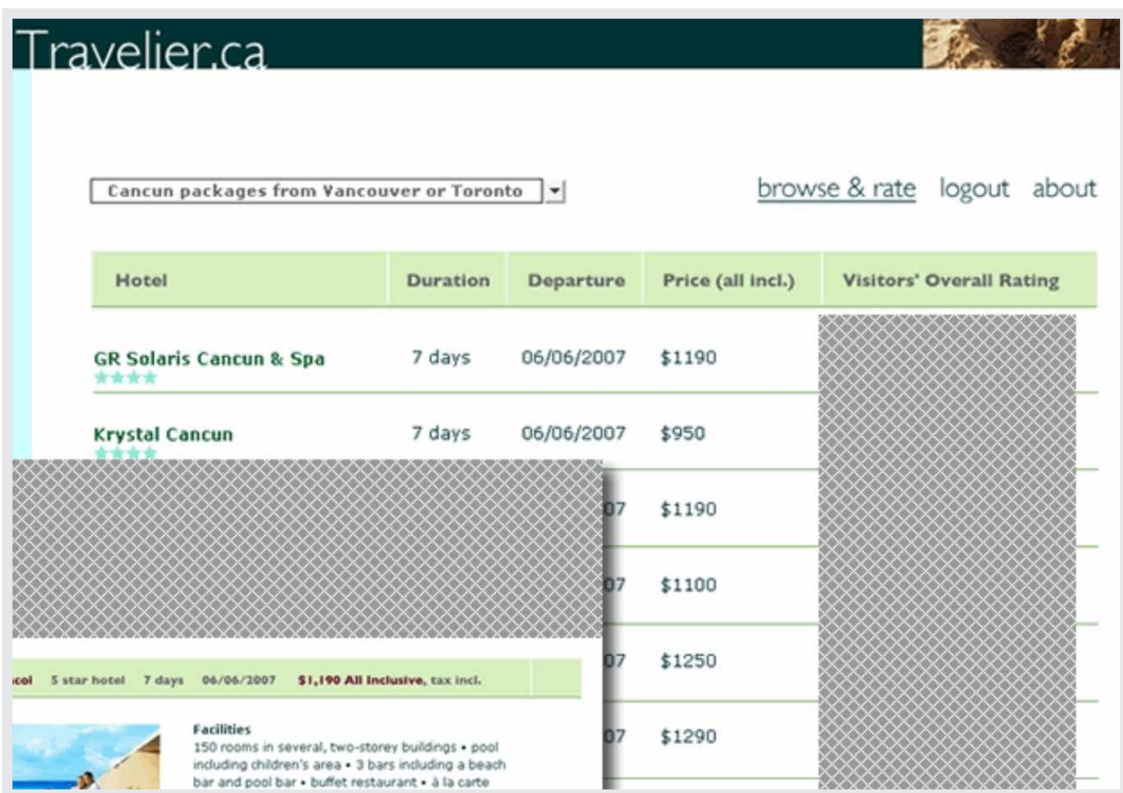
For many years involvement has been the object of considerable consumer-oriented research. Although there have been numerous definitions of involvement, Koufaris 2002(p. 211) summarized involvement as: "(a) a person's motivational state (i.e. arousal, interest, drive) towards an object where (b) that motivational state is activated by the relevance or importance of the object in question..." If the website permits involvement for the user, then it will result in an affective response that will be greater than an elicited cognitive reaction (Fortin and Dholakia, 2003). Online, involvement implies a user emotional response that includes absorption and excitement with website characteristics (e.g. Kumar and Benbasat, 2002; Santosa et al., 2005; Singh et al., 2003), and therefore encompasses elements of "flow". Jiang et al. (2010) refer to "affective involvement" as a heightened emotional feeling associated with a website comprising how users feel toward the website.

In terms of website antecedents that result in involvement, website interactivity has played a prominent role (Fortin and Dholakia, 2003; Johnson et al., 2006).[7] Interactivity potentially enables the user to have augmented control of the content, and thus offers an opportunity for the user to interact with the advertiser and/or other consumers (Fortin and Dholakia, 2003). Involvement was found to have a "pivotal role" (ibid, p. 394) in understanding how consumers interact with websites, and thus influences consumer loyalty toward the site. In research that examined different levels of interactivity in a fictitious vacation website (Cyr et al., 2009), five different web-poll designs[8] (Figure 40.8) were tested with users. The designs range from no user interaction in Treatment 1 to high interactivity and visualization capability in Treatments 4 and 5.

7. Specific definitions as to what interactivity represents vary (Johnson et al., 2006; Lee, 2005) but usually include elements such as user control of information display and content, website responsiveness to user queries, personalization tailored to unique users, and a feeling of connectedness often with other users (Dholakia et al., 2001; Lee, 2005). Cyr et al. (2009) defined perceived interactivity as allowing the user to control and access information on the site in a variety of ways, which is both personal and responsive.

8. 'For additional details on web-polls and their design refer to Cyr et al. (2009) or Ivanov et al. (2006).

Based on survey results, perceived website interactivity resulted in user perceptions of efficiency, effectiveness, enjoyment and trust, and ultimately online loyalty (Cyr et al., 2009). Although there were no statistically different results for the different web-poll treatments in Figure 40.8, additional qualitative analysis revealed that users had more positive impressions of the more interactive websites. Relevant to the topic of emotion and website design, different concepts emerged, which included "Affective" and "Aesthetic" categories. For instance, for Aesthetics users noted the websites were "visually appealing", "unique", "creative", "stylish and innovative". For the Affective category, users described the sites as "exciting", "makes the customers feel more empowered by allowing them to influence others using the poll system", "has a warm feeling to it", is "entertaining" and "fun".



The content above was common to all treatments. There were three pages in total: the front page with a list of hotels and a summary, and two detail pages. These two (one shown as the smaller cutout) included photos, description and the actual web-poll at the top. The gray-striped area is where design varied.




Treatment 1: Control version. No user interaction with web-poll; static indicator of other users' rating only.



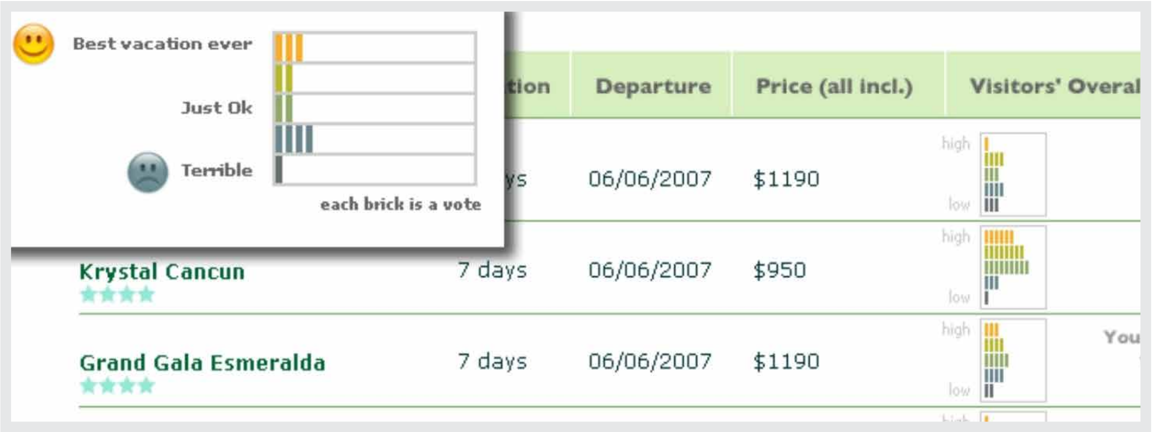

Treatment 2: Basic web-poll with conventional interaction (radio button) and simple information visualization (bar chart).




Treatment 3: Metaphor-rich web-poll. Cursor reveals foot icon across sandbox to select one of nine possible value combinations on a grid. Mini plot on front page with size of dot displaying number of votes.




Treatment 4: Flash version for enhanced user control. Cursor changes into foot icon, moving on scale continuously. Front page summary uses color lightness to represent weight. Bar levels give a positive/negative 'slope' for before and after.






Treatment 5: Enhanced Bar Chart version for visualizing user contribution. Users 'viscerally' plot their vote to the stack by adding a 'brick'

Figure 40.8: Five Levels of Website Interactivity (in Cyr et al., 2009).

### 40.2.3 Trust

Website usability can significantly impact trust (Flavián et al., 2006). In online environments numerous researchers have endeavored to understand the complexities inherent in trust (Bhattacherjee, 2002; Chen and Dhillon, 2003; Cheung and Lee, 2006; Gefen, 2000; Gefen et al, 2003;Jarvenpaa et al., 2000 ; Komiak and Benbasat, 2004; Koufaris and Hampton-Sosa, 2004;Rattanawicha and Esichaikul, 2005 ; Wang and Benbasat, 2005; Yoon, 2002)[9]. Online trust relates to consumer confidence in a website and willingness to rely on the vendor in conditions where the consumer may be vulnerable to the seller (Jarvenpaa et

9. Website trust has received ongoing attention concerning website design, and the literature in this area is vast. For an overview of online trust refer to Bhattacherjee, 2002; Corritore et al., 2003; Gefen et al., 2003; Jarvenpaa et al.1999, or McKnight et al., 2002. A useful outline of trust and dimensions of trust appears in Gefen and Straub, 2004. Closely aligned to trust is credibility. Considerable work on this latter topic has been done by Fogg and his colleagues, and a summary report appears here Also refer to the full report titled "How Do People Evaluate a Web Site's Credibility? At: http://www.consumerwebwatch.org/pdfs/stanfordPTL.pdf

al., 1999). Trust has an emotional component, and according to Komiak and Benbasat (Komiak and Benbasat 2006, p. 943) "[E]motional trust is defined as the extent to which one feels secure and comfortable about relying on the trustee". Unlike the vendor-shopper relationship established in traditional retail settings when trust is assessed in a direct and personal encounter, the primary communication interface with the vendor is an information technology artifact, the website. An absence of trust is one of the most frequently cited reasons that consumers refrain from purchasing from Internet vendors (Grabner-Kräuter and Kaluscha, 2003).

Various studies have shown that website trust is fundamental to e-loyalty including online purchase intentions (Flavián et al., 2006; Gefen, 2000; McKnight et al, 2004), and willingness by consumers to buy from an online vendor (Flavián et al., 2006; Laurn and Lin, 2003; Pavlou, 2003). Antecedents to user trust in websites vary, and include website design characteristics (Flavián et al., 2006) and design credibility (Green and Pearson, 2011). Everard and Galletta (2005-6) conducted a study in which they examined how presentation flaws on websites (e.g. errors, poor style, incompleteness) influence user perceptions. More specifically, websites were experimentally created to demonstrate good versus poor quality (Figure 40.9). Results of the investigation found that user perceptions of flaws on the website related to perceived quality—which was in turn directly related to trust and intention to purchase from the store. Thus, careful attention to detail and the elimination of design flaws has a positive impact on the user.

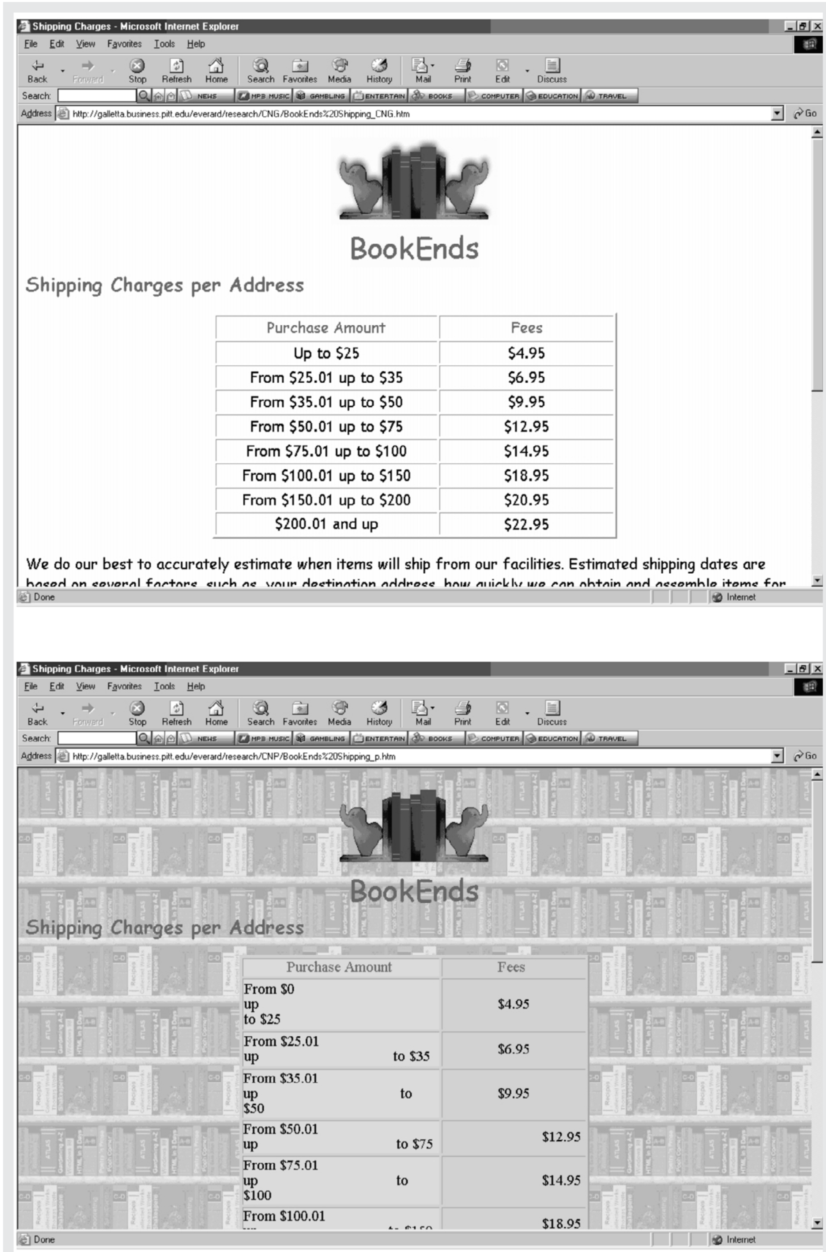





FIGURE 40.9: Samples of Web Pages for Good (upper) versus Poor Style (in Everard and Galletta (2005-6).

A relationship exists between beauty of a website and trust (Karvonen, 2000). More specifically, images have the power to enhance consumer trust in a vendor. In this vein, jewelry retailer Tiffany and Co. invested in digital imaging technology to ensure images of jewelry are presented on its website in such a way as to instill trust in potential buyers (Srinivasan et al., 2002). Refer to Figure 40.10.

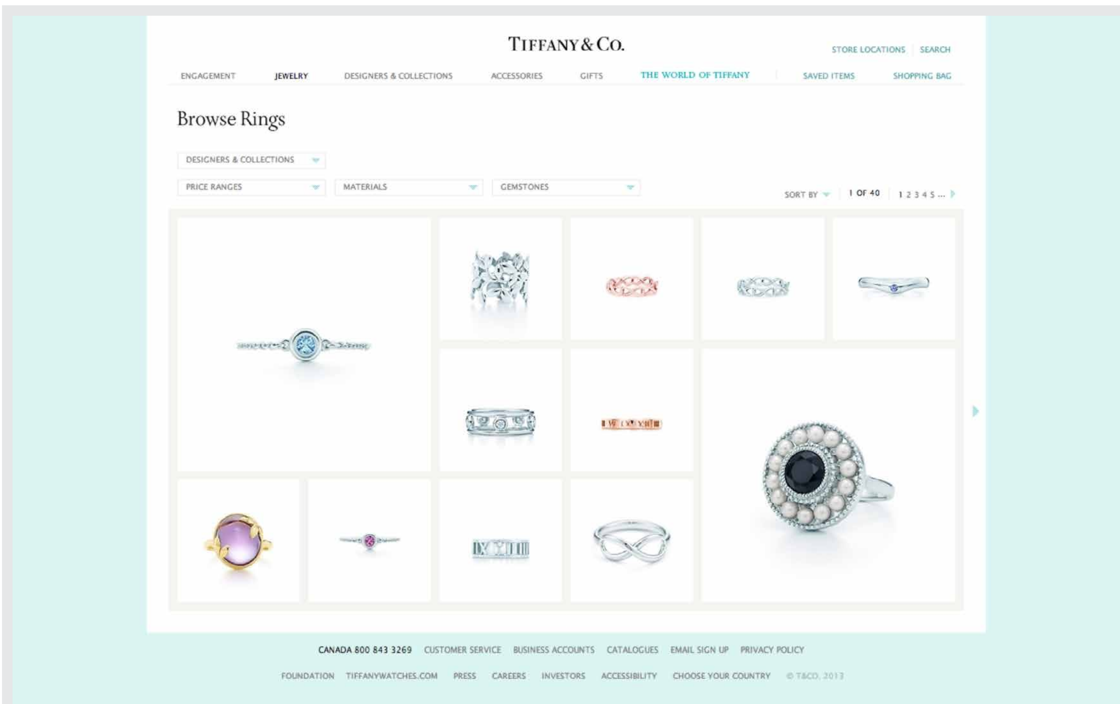





Figure 40.10: A Sample Web page from Tiffany and Co.

Additional triggers of online trust include vendor size (van der Heijden et al., 2000), perceived vendor reputation (Jarvenpaa et al., 2000; Koufaris and Hampton-Sosa, 2004;Van der Heijden et al., 2000), service quality (Gefen, 2002), social presence (Gefen and Straub, 2003), and perceived security control (Koufaris and Hampton-Sosa, 2004).

### 40.2.4 Satisfaction

Satisfaction on the web relates to "stickiness" and the sum of all the website qualities that induce visitors to remain at the website rather than move to another site (Hoffman and Novak, 1996). An effectively designed website engages and attracts online consumers, resulting in online satisfaction (Agarwal and Venkatesh, 2002; Fogg et al., 2002; Hoffman and Novak, 1996; Koufaris, 2002; Lindgaard and Dudek, 2003;Nielsen, 2001 ; Palmer, 2002; Szymanski and Hise, 2000; Yoon, 2002).

Elements of website design that contribute to satisfaction are numerous and varied. Palmer (2002) validated design metrics for websites and found that site organization, information content, and navigation are important to website success—including user intent to return to the site. In other research, website design and the "ambience associated with the site itself and how it functions" is an antecedent to satisfaction (Straub, 1989). In alignment with other researchers (e.g. Agarwal and Venkatesh, 2002; Cronbach, 1971; Falk and Miller, 1992), website satisfaction is defined here as overall contentment with the online experience. Website satisfaction is frequently a predictor of e-loyalty (Flavián et al, 2006; Kim and Benbasat, 2006; Lam et al., 2004; Laurn and Lin 2003; Yoon, 2002).

### 40.2.5 Summary

From the preceding, it is clear that if websites are effective and are able to arouse responses in users such as enjoyment, involvement, trust, or satisfaction, then they will be successful in enticing users to return to the site. These findings are intuitive, but in the present article they are also founded on considerable systematic and rigorous investigation by researchers—mostly in the information systems area. However, beyond pure research, these results have merit for practitioners such as web strategists and designers. As a powerful communication mechanism for commercial or other use, effective website design has the ability to persuade. Hence, a worthy goal is the amalgamation of work from both the academic research and design communities to forge new and deeper understandings as to how emotion and websites are tied together. This calls for integrated and multidisciplinary approaches—as well as multiple methods for assessing user reactions to website design elements.

## 40.3 GRAPHICAL WEBSITE DESIGN ELEMENTS: A FOCUS ON COLOR AND IMAGES

Although a number of website elements could be considered in depth concerning their ability to impact the user, this chapter focuses on the website characteristics of color and imagery. These two elements are not only central in website design, but they also have interesting cultural implications for the user that merit future consideration. Thus we begin with an overview of color and website design—including how color preferences vary by country. This is followed by an overview of images, similarly with some discussion as to how imagery is perceived by users from different countries.

### 40.3.1 Color and Emotion

The study of color and color appeal has interdisciplinary connections. Color has been studied by a variety of researchers—from artists to zoologists—including the use of color in art (Fornell, 1982) or visual perception (Gorn et al., 1997). Psychologists have been interested in the effect of color on individual preferences (Goldberg et al., 2002). Some colors are able to arouse and excite an individual, while other colors elicit relaxation. Research on color suggests hue (as in primary colors red, blue, yellow), brightness (light colors such as white versus dark colors such as black or gray), and saturation (intense versions of a color versus pastels) all have an effect on individual reactions and perceptions (Latomia and Happ, 1987).

"Colors are known to possess emotional and psychological properties" (Lichtle, 2007, p. 91), and have the potential to convey commercial meaning in products, services, packaging, and Internet design. For many years, marketers have utilized the power of color for logos or displays to build consumer confidence in the corporate brand (Lui et al., 2004; Rivard and Huff, 1988). Cool colors such as blue and green are generally viewed more favorably than warm colors such as yellow or red (Goldberg and Kotval, 1999; Latomia and Happ, 1987; Marcus and Gould, 2000). Blue is generally associated with "wealth, trust, and security" (Lichtle, 2007) and is universally liked (Carte and Russell, 2003; Meyers-Levy and Peracchio, 1995; Nielsen and Del Galdo, 1996). In part, this explains the use of blue by corporate entities such as banks (at least in North America) or IBM to establish a professional and credible image. Red is the color symbolizing Coca Cola. Alternately, orange denotes "cheapness" (Lichtle, 2007).

Although color has the potential to elicit emotions or behaviors, in only a handful of studies are various website color treatments empirically tested regarding their impact on user trust or satisfaction. However, Cyr (2008) found that visual design of the website (which includes color) resulted in trust, satisfaction, and loyalty. Further, Kim and Moon (1998) examined color and four other design factors on an online banking interface. Color elements examined were color tone (e.g. warm or cool), main color (e.g. primary or pastel), background color, brightness, and symmetry (how color was organized). The findings show that color has a main effect on trustworthiness of the interface. Color likewise has an influence on behavioral intentions such as customer loyalty, with blue producing stronger buying intentions than red (Becker, 2002; Latomia and Happ, 1987).

In a study aimed to investigate the impact of color related to user emotion and perceptions of website appeal, Bonnardel et al. (2011) tested a variety of colors to determine whether the user found the site to be pleasing, appealing, and appropriate (Figure 40.11). By varying hue and intensity 23 pages were created.[10] Testing with groups of participants with varied backgrounds, including web designers,

10. Designers and researchers are referred to the analysis sections of the paper by Bonnardel et al., (2011) for a useful elaboration as to how color was measured using hierarchical cluster analysis based on color hue values.

yielded consistent color effects—with blue as the most preferred color, and gray as the least preferred. Additionally, color impacted how users navigated the site and the information they retained.




Figure 40.11: Samples of Diverse Color Use on Websites (in Bonnardel at al., 2011).

## 40.3.2 Culture and Color

Two opposing views exist when culture, color, and cognition intersect. A Universalistic view prescribes generic cognitive processing of color perception. Alternately, Cultural Relativism suggests that color perception is mostly shaped by culturally specific language associations and perceptual learning (Berlin and Kay, 1969; Kay et al., 1991). For the Internet, the prevailing view is more aligned to Cultural Relativism in that ideally websites should be developed for specific cultural and user groups—termed website localization (Barber and Badre, 1998; Cyr and Trevor-Smith, 2004). When localizing a website, in addition to language translation, details such as currency, color sensitivities, product or service names, images, gender roles, and geographic examples are considered[11]. Localized website design creates alignment to user expectations, and according to an ISO quality standard this is a critical success factor for effective and efficient task completion.

11. While an overview of localization procedures is beyond this article, refer to "A Guide to Localization Management" which includes useful practices, including checklists. Accessed 13th March 2013 here.

Color connotes different meaning in different cultures (Bagozzi and Yi, 1989; Barber and Badre, 1998; Singh et al., 2003). Red means happiness in China, but danger in the United States. In a cross-cultural study focused on understanding the meaning and preferences for ten colors across eight countries, blue was generally considered "peaceful" or "calming" (Madden et al., 2000). In contrast, brown and black had associations of "sad" and "stale" across cultures. Other colors, such as yellow and orange, showed less cross-cultural consistency in terms of how they were perceived.

In the information systems and e-commerce realm, a growing number of studies have been published with respect to culture and website design, and the subsequent effect on users. In a report in which 27 research studies published in 16 different journals were evaluated for website cultural congruency (e.g. cultural adaption), strong empirical support was provided for the positive impact of cultural congruency on performance measures including website effectiveness (Vyncke and Brengman, 2010).[12] Other investigations support unique user preferences for website design characteristics in different countries and cultures. For instance, for a study in which domestic and Chinese websites for 40 American-based companies were systematically compared, significant cultural differences were uncovered for all major categories tested (Singh et al., 2003).[13]

Especially rare are rigorous studies which are primarily focused on color in website design across cultures. This area represents an important and overlooked

12. Designers and researchers are referred to the analysis sections of the paper by Bonnardel et al., (2011) for a useful elaboration as to how color was measured using hierarchical cluster analysis based on color hue values.

13. Refer to Singh et al. (2009) for a diagnostic framework to analyze the degree to which a company localizes website content resulting in overall quality of the localization effort. A scorecard is provided for practitioners to use in the assessment of website localization with 12 categories such as content depth and synchronization, navigation, web page structure, graphics, colors, translation quality, among others. Also of interest may be Smith et al. (2004), who developed a practical "process model" for developing usable cross-cultural websites. This includes: conducting an audit of websites; developing a "cultural fingerprint" to determine cultural needs of users; and how to deal with issues associated with user evaluation, as well as a cross-cultural development team.

topic. According to Noiwan and Norcio (Noiwan and Norcio 2006, pp. 103, 104), "[E]mpirical investigations on the impacts of cultural factors on interface design are absolutely vital… Interface designers need to understand color appreciation and color responses of people in different cultures and regions."

In a study comparing Indian and U.S. websites related to language, pictures, symbols, and colors substantial differences were found regarding the use of color (Kulkarni et al., 2012). The Indian portal (Figure 40.12) had multiple colors on a white background, while the U.S. website (Figure 40.13) hosted only blue and red, which, as the authors point out, are represented on the national flag.




Figure 40.12: An Indian Portal (in Kulkarni et al., 2012).


*Courtesy of USA.gov. Copyright: pd (Public Domain (information that is common property and contains no original authorship)).*

FIGURE 40.13: A U.S. Portal (in Kulkarni et al., 2012).

From this study it appears that the use of specific colors on websites in different countries can also impact color appeal, trust and satisfaction.

Cyr and Trevor-Smith (2004) examined design elements, including the use of color, using 30 municipal websites in each of Germany, Japan, and the U.S (90 websites in total). Use of symbols and graphics, color preferences, site features (links, maps, search functions, page layout), language and content were examined, and significant differences were determined in each website design category. Colors used on a website were matched to a color wheel and assigned a numerical value by independent raters based on the percentage of the page on which a color appears.[14] Fifteen colors were used across the websites. Relevant to the countries

14. The Cool Ruler application tool was used and allows researchers to measure the page and the sections of

in this investigation, blue was most popular on German websites, while gray was the color most often appearing on American websites. Japanese are known to prefer brighter colors such as yellow (also supported by Noiwan and Norcio, 2006).

Building on empirical results from the few studies in which color is examined on websites, Cyr et al. (2010) examined the influence of blue, gray, and yellow on color appeal, trust, satisfaction, and online loyalty. Figure 40.14 illustrates a sample of how color is adapted.[15] Testing was conducted on 30 participants from each of Canada, Germany, and Japan (90 total). These countries were chosen based on known cultural diversity (e.g. Hofstsede, 1980). Data was collected using a questionnaire, interviews, and an eye-tracker.[16] Website color appeal was found to be a significant determinant of website trust and satisfaction, with differences across cultures. More specifically, Canadians, Germans, and Japanese all tend to dislike yellow on websites. Users in all three countries most prefer the blue sites, contrary to expectations that Japanese would prefer a bright color.

Of particular relevance to a discussion of emotion and websites, from interview analyses five concepts emerged from the data related to use of color on the website: (1) aesthetics concerning artistic appeal, (2) affective or emotional quality of the color, (3) functional quality, (4) harmony (i.e. balance), and (5) appropriateness (Cyr et al., 2010). As already noted, color has the ability to elicit an emotional or affective response toward a website, and this varied across the various groups. For instance, in all three countries blue had positive affective quality while in none of the countries did users mention this was the case for gray.

---

the page. The section of the page with a certain color is then divided by the total page size to determine the percentage of color on the page. The ruler appears on the screen and can measure in inches, pixels, and centimeters. Measurements are made by adjusting the ruler to either a horizontal or vertical orientation and then using the ruler as if on a sheet of paper. The Cool Ruler application is available at: http://www.softpedia.com/get/Desktop-Enhancements/Other-Desktop-Enhancements/Cool-Ruler.shtml.

15. Note these are not the exact versions of the websites that appeared in the research by Cyr et al. (2010), and the viewing patterns have been added.

16. Although an eye-tracking devise to track a user's eye movements and fixation points has been used in research for some time, its use is recent in the information system area, and specifically to examine affective reactions in e-commerce. For further details about use of an eye-tracker refer to Cyr et al. (2009) or Djamasbi et al. (2010).

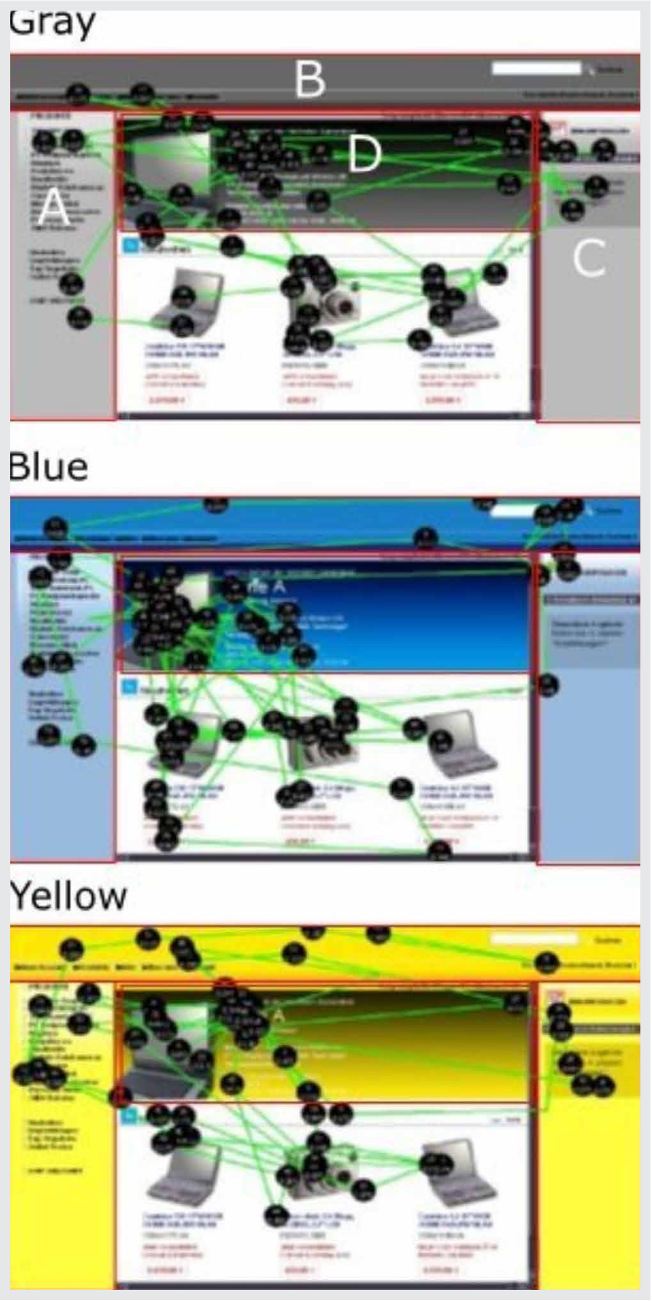





FIGURE 40.14: Sample Websites indicating Color Zone Treatments and Look Zones for a German Website (preliminary experimental treatments adapted from Cyr et al., 2010).

### 40.3.3 Imagery and Emotion

Image design for websites may include elements of balance, emotional appeal, aesthetics, and uniformity of the overall graphical look of the site. This encompasses web elements such as use of photographs, colors, shapes, or font type (Garrett, 2003). The aesthetics of website design are considered related to the "overall enjoyable user experience" (Tarasewich, 2003, p. 12). The use of photographs in websites has been debated among usability experts in a discussion of whether photographs unnecessarily clutter up the website, slow it down, and disrupt its functionality (Riegelsberger, 2002). Alternately, images have been found to attract viewer attention (Riegelsberger, 2002), and increase credibility (Fogg et al., 2002).

In advertising, images are used to convey product and brand information—and to elicit emotional responses from consumers (Branthwaite, 2002; Kamp and MacInnis, 1995; Swinyard, 1993). In one study, a picture of a spray bottle of window cleaner composed of purple berries aligned with the verbal statement "bring home a fresh fruit orchard" generated "positive inferences" from viewers (Phillips and McQuarrie, 2005). In other research, happy or angry faces were flashed on a screen while people examined Chinese ideographs. The type of face affected "liking ratings" of the ideographs. Even small alterations in an image can impact product evaluations. For instance, changing the camera angle of a product can influence the viewer's attitude toward the product (Meyers-Levy and Peracchio, 1992).

Online, the visual design of an e-commerce website is important because it improves website aesthetics and emotional appeal (Garrett, 2003; Liu et al., 2001; Park et al., 2005), which may in turn lead to more positive attitudes toward an online store (Fiore et al., 2005).Loiacono et al. (2007) created an instrument to measure consumer evaluation of websites, and found that visual appeal and consistent images resulted in entertainment for the user, ultimately leading to intentions to reuse the site in the future. In research on banner ads, different formats using either text or images were manipulated. Viewers consistently rated the versions with images as more positive and effective (Yoon, 2002).

Age makes a difference as to how images are processed and appreciated. For example, some studies suggest that web aesthetics and visual design may be especially important to Generation Y users (e.g. born 1977 to 1990) (Tractinsky, 2004; 2006). Other studies have shown that visual appeal of the homepage for an online vendor impacts impressions of vendor image and merchandise quality for Generation Y users (Oh et al., 2008). Users younger than age 25 seek fun when shopping online, and respond positively to personalized product offers, custom-designing products, and seeing the profiles of previous customers who purchased an item. In a two pronged study that included surveys and investigations where participants looked at a web page using an eye-tracking device, Generation Y users exhibited specific preferences for a large main page, images of celebrities, little text, and a search feature (Djamasbi et al., 2010). In research that compared Generation Y users and Baby Boomers (e.g. born 1946 to 1964) using an eye-tracking device and self-report measures, both generations reported similar aesthetic preferences, and liked pages with images and little text (Djamaski et al., 2011). However, viewing patterns differed, with Baby Boomers having significantly more eye fixations that covered more of the pages.

### 40.3.4 Human Images and Image Appeal

Online, images of people are used to induce favorable emotional responses (Riegelsberger et al., 2003) and to draw attention (Tullis et al., 2009). Adding images of players engaged in an online text chatting game increased cooperation (Zheng et al., 2002). The use of human images is thought to increase the website's aesthetics and playfulness—and therefore positively influence the user (Liu et al., 2001).

For many years advertising has relied on imagery using "friendly faces" to build a positive attitude toward products (Giddens, 1990), as shown with the use of faces on sample banner ads, lifestyle photos, and opinion articles in Figure 40.15.


*Courtesy of Various Copyright Holders (synthesized by Tullis et al., 2009). Copyright: compositeWork-WithMultipleCopyrightTerms (Work that is derived from or composed of multiple works with varying copyright terms and/or copyright holders).*

Figure 40.15: Sample Use of Faces (in Tullis et al., 2009).

Online trust can be established through virtual re-embedding of content and social cues (Riegelsberger et al., 2003; 2005). In a study using pages from the online shop of a well-known British supermarket chain, identical pages were created with one exception: one page contained a photograph of a human face while another had a box of text the identical size as the photograph. Viewers were more attracted to the photograph, leading to the conclusion that "the face is a very important source of socio-emotional cues....Advertisers have found that photographs of faces attract attention and create an immediate affective response that is less open to critical reflection than text we read" (Riegelsberger, 2002, p. 1). For an online banking website, inclusion of employee photographs resulted in attributions of trustworthiness (Steinbrück et al., 2002). A photograph of the author in an online magazine article resulted in greater perceived trustworthiness of the article (Fogg et al., 2001).

Using an eye-tracking device, Djamasbi et al. (2012) examined the impact of facial images on viewing behavior, and the number of user eye fixations on web pages[17]. Using the theory of visual hierarchy (e.g. the order in which information is communicated to users) objects on the website were manipulated related to the presence or absence of facial images, and their location either above or below the mid-point on the page. Typically, users follow an F-shaped pattern when viewing web pages—that is, they look along the left hand portion of the page, and particularly the top left hand area (Buscher at el., 2009). Figure 40.16 shows the pages used in Djamasbi's study.

17. Details regarding how viewer behavior was tracked and analyzed can be found in the paper. This includes the use of heat maps which indicate fixation patterns.


*Courtesy of Unknown Website (manipulated in Djamasbi et al., 2012). Copyright: pd (Public Domain (information that is common property and contains no original authorship)).*

Figure 40.16: Web Pages with Faces and their Location (in Djamasbi et al., 2012).

Findings from the above study (Djamasbi et al., 2012) revealed that faces did not necessarily increase viewing time or the number of people who viewed the areas with images. However fixation patterns in these areas were affected. For instance, viewing was more dispersed when faces are present - and thus users scanned faces, titles, and text (Figure 40.17, a and b. Also, faces above the mid-point on the page attracted significantly longer fixations, which negatively affected user performance on a performance task since users diverted their attention from key information such as titles.




Figure 40.17: Heat Maps when Browsing Faces or Text (in Djamasbi et al., 2012).

### 40.3.5 Images and Culture

Cultural differences exist with respect to the use of images. In collectivist cultures such as China or Japan, users have a strong preference for visuals (Szymanski and Hise, 2000) including pictures, bright colors, and animation (Cyr et al., 2005). Design elements including Visual Design influence user trust, which also varies by country. In an e-commerce investigation in Canada, Germany, and China, Visual Design only resulted in trust for Chinese users (Cyr, 2008). In a study of Indian websites, images were found to be important carriers of culture, and most websites have pictures of Indian people and culture (Kulkarni at el., 2012). However, as the researchers note (Ibid), globalization is changing the type of imagery used to be reflective of Western culture (Figure 40.18).




Figure 40.18: Westernization of an Indian Website (Kulkarni et al., 2012).

Other research examined the relationship of human images to image appeal, perceived social presence (e.g. the warmth and sociability of the website), and trust (Cyr et al., 2009). A controlled experiment was conducted with different levels of human images represented on a web page: human images with facial features, human images without facial features, and no human images (Figure 40.19).

Data was collected on websites for Canada, Germany, and Japan using a questionnaire, interviews, and an eye-tracking device (Cyr at al., 2009). The model was supported in that human images with facial features resulted in user perceptions that the website was more appealing, had warmth or social presence, and was more trustworthy. Of interest, participants in all countries spent the most time viewing the partial image (e.g. the condition with human images but with no facial features) which they described as “unnatural” and “distracting”. Additional analyses revealed subtle differences in the perception of human images across the three countries. While the general impact of human images appears universal across country groups, based on interview data four concepts emerged: aesthetics (visual qualities such as pretty or bright); symbolism (implied meaning — for instance an image of a man and little girl may be a representation of father and daughter); affective property (and eliciting emotion); and functional property (structural properties such as information design, navigation, layout). Based on interview data, participants from each culture focused on different concepts based on both positive and negative assessments (Figure 40.20).


My Order | View Cart
Glossary
Product Registration
Customer Support
Francais
Search
All Products
Go
Product Information
Accessory Store
VAIO Computers
CLIÉ Handhelds
Digital Cameras
Camcorders
Digital Music Players
Portable Audio
Televisions
DVD/VCR
Home Audio
Car Audio
Memory Stick
Accessory Store
Shopping Support
Member Sign In | Member Registration
New 5-Megapixel Cyber-shot cameras from Sony!
What's your Style?
Cyber-shot
DSC-W1
DSC-P93
BONUS Cyber-shot® digital camera
with VAIO® V
VAIO
Save $200 on VAIO




My Order | View Cart
Glossary
Product Registration
Customer Support
Francais
Search
All Products
Go
Product Information
Accessory Store
VAIO Computers
CLIÉ Handhelds
Digital Cameras
Camcorders
Digital Music Players
Portable Audio
Televisions
DVD/VCR
Home Audio
Car Audio
Memory Stick
Accessory Store
Shopping Support
Member Sign In | Member Registration
New 5-Megapixel Cyber-shot cameras from Sony!
What's your Style?
Cyber-shot
DSC-W1
DSC-P93
BONUS Cyber-shot® digital camera
with VAIO® V
VAIO
Save $200 on VAIO




Figure 40.19 A-B-C: Three Levels of Human Images on a Canadian Website (in Cyr et al., 2009) (Top - No Human Images, Middle — Human Images but No Facial Features, Bottom — Facial Images.

**Table 10. Summary of Interview Analysis Across Cultures**

| Categories | Canada | Germany | Japan |
|---|---|---|---|
| High-Human images | Positive: Aesthetic, Affective | Positive: Functional, Affective, Symbolic | Positive: Affective, Symbolic |
| | Negative: Aesthetic, Functional | Negative: Functional | Negative: None |
| Medium-Human images | Positive: Symbolic | Positive: Functional | Positive: Affective, Symbolic |
| | Negative: Aesthetic, Functional | Negative: Affective | Negative: Aesthetic |
| No-Human images | Positive: Functional | Positive: Functional | Positive: None |
| | Negative: Aesthetic, Affective | Negative: Affective | Negative: Affective |



Figure 40.20: Summary of Emerging Concepts by Culture (in Cyr et al., 2009).

It is also interesting to reflect on the interview comments from participants in this study, and how the various images impacted them (Cyr at al., 2009). Overall, users in all countries liked the human images with facial features. It was noted: "Pictures of people add a human level and we can relate to them" (Canada); "It's important to show images with people, their smiles, their emotions" (Germany); "The sites with the images of people looked warm and trustworthy" (Japan). In contrast, reactions to the no-human image condition are largely negative. Participants across the three countries criticized the blank images, and perceived the website as unfriendly. According to one Canadian: "When the site just had the actual product and absolutely no human images, it did make it look stark." Alternately, a few Canadians and Germans perceive websites with no human pictures as functional. That is, the site is not cluttered with distracting pictures (Cyr at al., 2009).

Between countries, Canadians tended to focus on aesthetic aspects of the images and website, while Germans and Japanese did not mention them. Germans commented on symbolism (i.e., community activities) of the images, as well as functional properties of the website. In the latter case, one German remarked, "I don't like pictures of people on the site. If I want to buy a computer, I don't need anything with human pictures on it. I just want to know the facts." Japanese participants mentioned they also like the community related symbolism of these images (Cyr at al., 2009).

### 40.3.6 Summary

The preceding evidence shines a light on the importance of color and image for eliciting user emotion on websites. While research, particularly usability studies, has proven this for some time, in the current context empirical linkages are established between color and images with specific user outcomes such as trust, satisfaction, and loyalty, among other reactions. Although research has generally considered the impact of design principles such as color or images, it is significant that considerably less attention has been paid to website design elements and their variance across cultures. This is the case, despite the importance of cultural sensitivity of website design. Thus, it would be beneficial if both researchers and practitioners were to investigate website localization and the impact it has on perceptions and emotions in diverse cultures. Other characteristics, such as age or gender likewise deserve attention concerning the use of color and imagery.

Finally, while most research relies on survey data to evaluate website design characteristics, studies which utilize multiple sources of data collection yield rich and interesting results. For instance, in some of the research discussed above, interview data and eye-tracking results provide nuanced information as to how users react to design elements. In addition to unconscious data provided by an eye-tracker, in recent studies physiological measures such using functional magnetic resonance imaging (fMRI) have been used to explore topics in the information systems domain. As one example, in a study measuring user brain activity while shopping on eBay using fMRI, differences were found between men and women concerning brain areas that encode trustworthiness (Riedl et al., 2010). [18]

## 40.4 SOCIAL ELEMENTS AND WEBSITE DESIGN

Building on the preceding, but broader than human imagery, there are a variety of socially oriented website elements which can elicit hedonic responses in users. Website social elements, and more specifically website social presence as mentioned above, is known to result in user enjoyment (Cyr et al., 2007; Hassanein and Head, 2007). In the following, perceived social presence is discussed, including differences between men and women regarding desired levels of social presence. In addition, cultural implications for social design elements are outlined, although little research has been conducted in this area.

18. Refer to Riedl et al., for details of how fMRI can be used to explore topics in e-commerce or other areas.

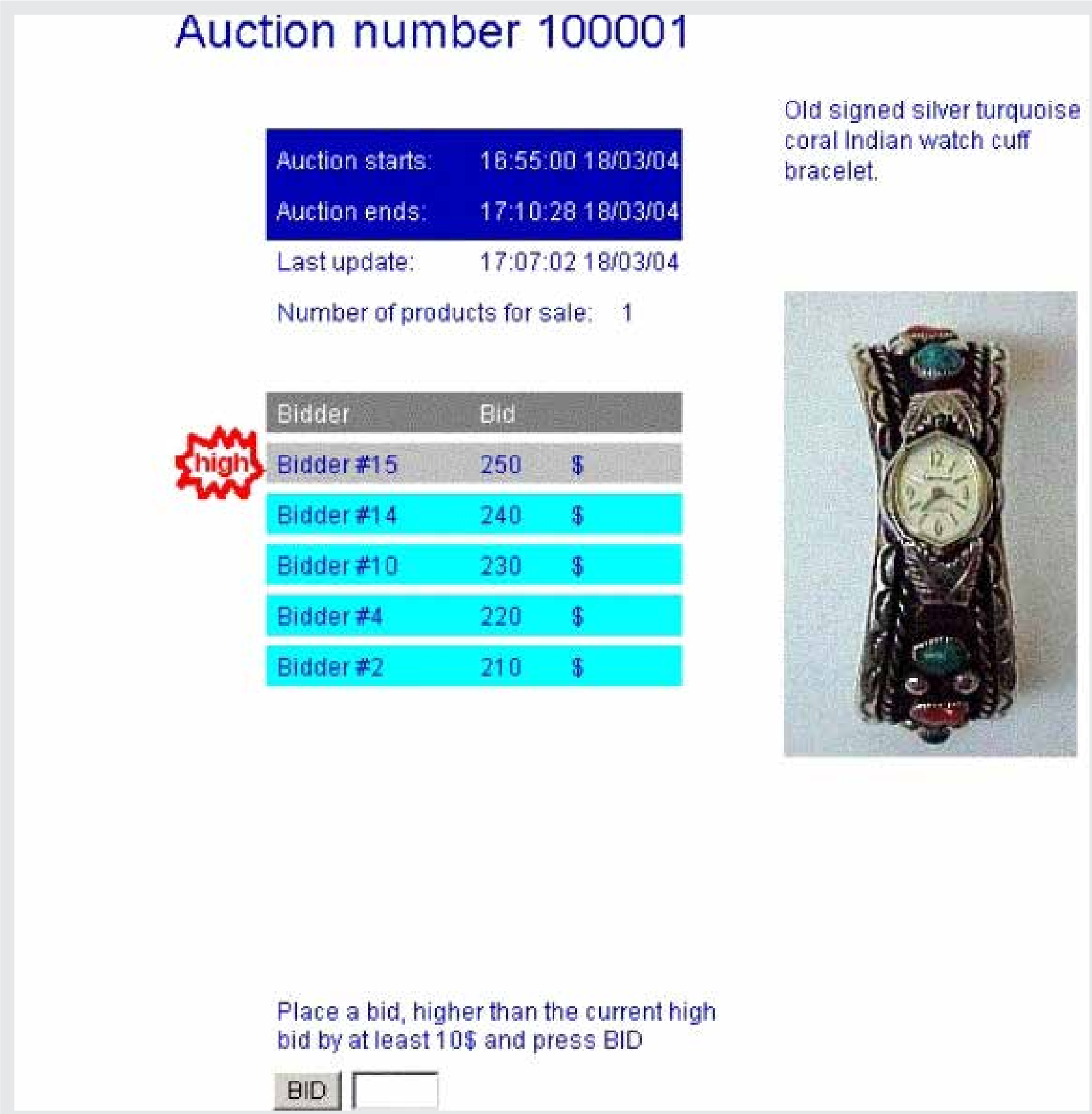
### 40.4.1 Perceived Social Presence

Perceived social presence has been defined as “the extent to which a medium allows users to experience others as being psychologically present” (Gefen and Straub, 2003, p. 11). Perceived social presence implies a psychological connection with the user who perceives the website as “warm”, personal, sociable, thus creating a feeling of human contact (Yoo and Alavi, 2001).[19] In one study, social presence has been segmented into “affective social presence” and “cognitive social presence” (Shen and Khalifa, 2009), although most research uses a uni-dimensional construct. Affective social presence refers to emotional responses aroused in the user by virtual social interaction. Cognitive social presence refers to the user’s belief regarding relationships with others in the social context.

Examples of website features that encourage social presence are socially-rich text content, personalized greetings (Gefen and Straub, 2003), human audio (Lombard and Ditton, 1997), or human video (Kumar and Benbasat, 2002). Gefen and Straub (2003) suggested that pictures and text are able to convey personal presence in the same manner as do personal photographs or letters. In addition to perceived social presence resulting in online enjoyment, social presence has implications for website involvement (Kumar and Benbasat, 2002; Witmer et al., 2005); website trust (Cyr et al., 2007; Gefen and Straub, 2003; Hassanein and Head, 2007); and utilitarian outcomes such as perceived usefulness (Hassanein and Head 2005-6; 2007).

In research on Internet auctions, two conditions of social influence were presented to participants: (1) interpersonal information in the form of text, or (2) “virtual presence” that included pictures of other bidder’s faces (Rafaeli and Noy, 2005) (Figure 40.21).

19. For a more complete review of definitions of social presence refer to Appendix II in Shen and Khalifa (2009)


Auction number 100001
Old signed silver turquoise coral Indian watch cuff bracelet.
Auction starts: 16:55:00 18/03/04
Auction ends: 17:10:28 18/03/04
Last update: 17:07:02 18/03/04
Number of products for sale: 1
Bidder Bid
high
Bidder #15 250 $
Bidder #14 240 $
Bidder #10 230 $
Bidder #4 220 $
Bidder #2 210 $
Place a bid, higher than the current high bid by at least 10$ and press BID
BID

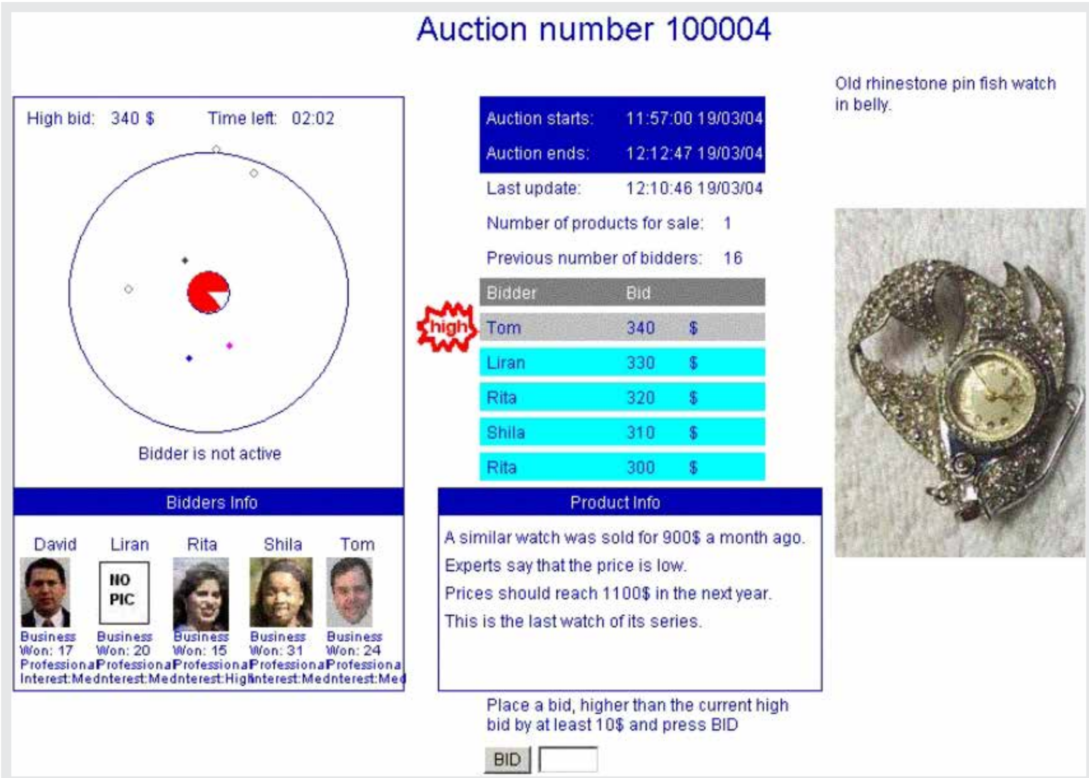


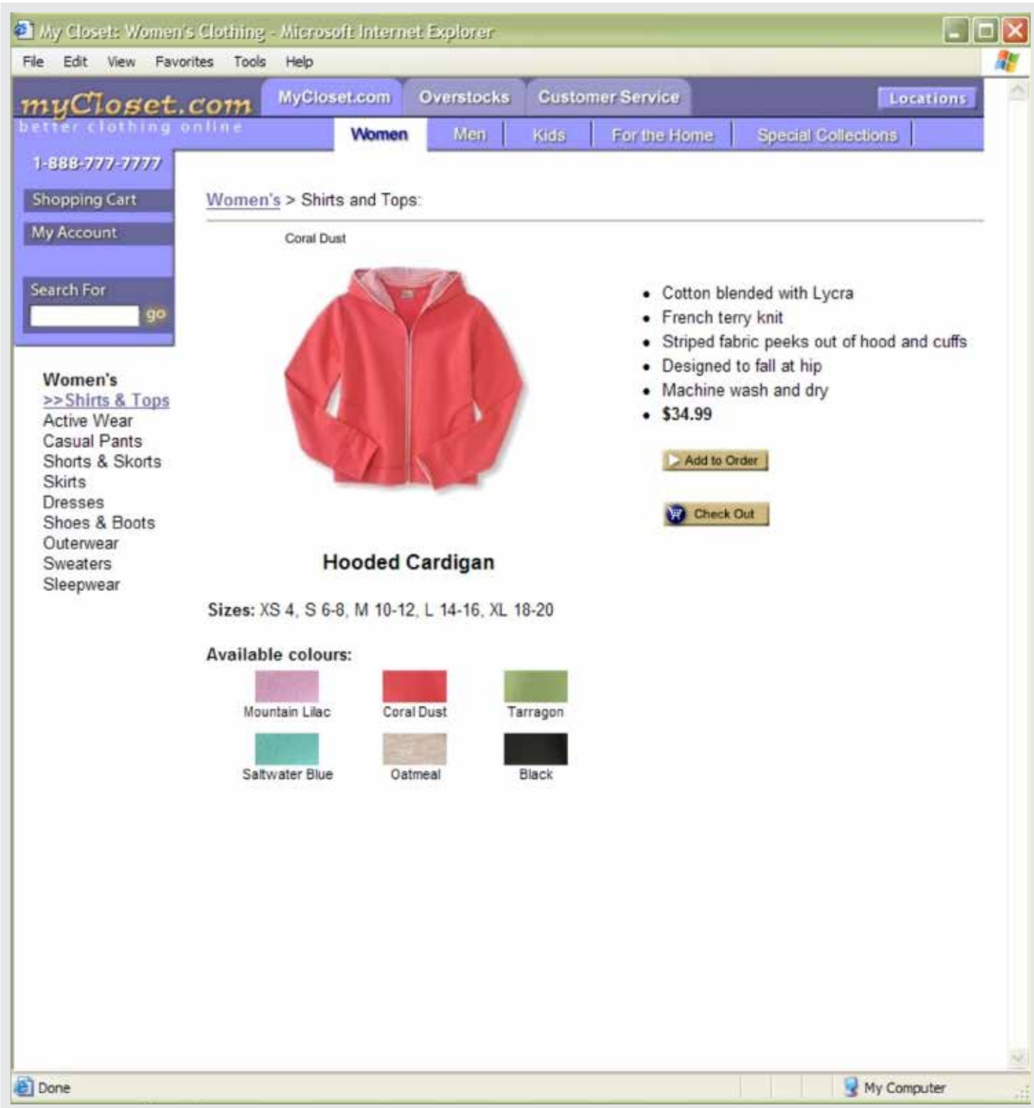


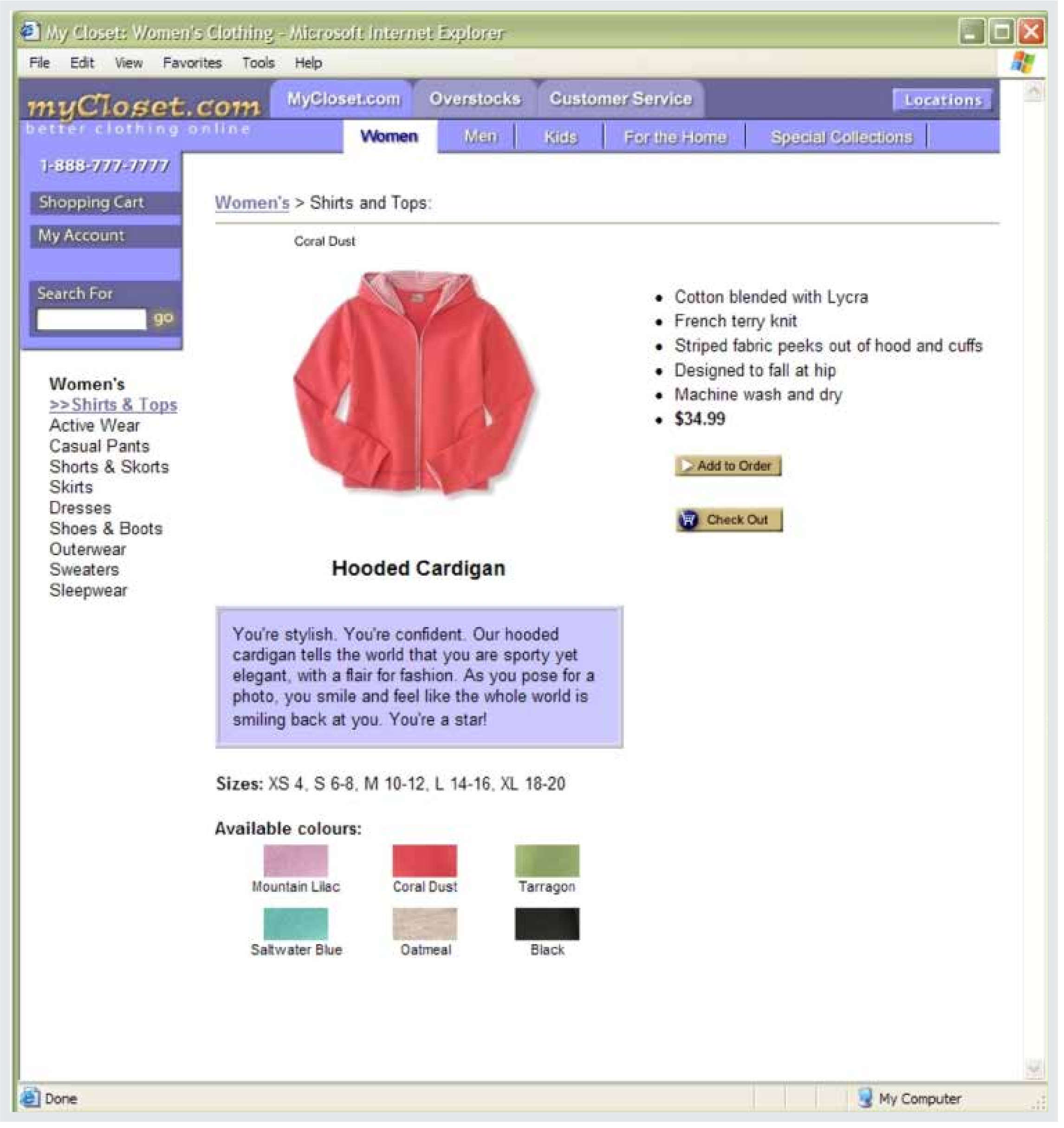

Figure 40.21 A-B: Difference Levels of Social Influence on Websites (in Rafaeli and Noy, 2005).

Results indicated the effect of interpersonal information on bidding behavior was not as important as the effect of virtual presence. The authors explained their results as being related to "the enthusiasm with facial cues [by users] and perception of other's presence" (Rafaeli and Noy, 2005, p. 172). The incorporation of human or human-like faces in online environments provides online participants with a stronger sense of community (Donath, 2001).

Previous research has manipulated Internet shopping conditions to investigate online social presence on an apparel website (Hassanein and Head, 2007).

In a low social presence condition functional text and a basic product picture appeared; in the medium condition a basic product picture appeared with emotive/descriptive text; and in the high social presence condition pictures depicted human figures interacting with the product as well as rich and emotive text (Figure 40.22).




My Closet: Women's Clothing - Microsoft Internet Explorer
File Edit View Favorites Tools Help
myCloset.com
better clothing online
MyCloset.com
Overstocks
Customer Service
Locations
Women
Men
Kids
For the Home
Special Collections
1-888-777-7777
Shopping Cart
My Account
Search For
go
Women's
>>Shirts & Tops
Active Wear
Casual Pants
Shorts & Skorts
Skirts
Dresses
Shoes & Boots
Outerwear
Sweaters
Sleepwear
Women's > Shirts and Tops:
Coral Dust
Cotton blended with Lycra
French terry knit
Striped fabric peeks out of hood and cuffs
Designed to fall at hip
Machine wash and dry
$34.99
Add to Order
Check Out
Hooded Cardigan
You're stylish. You're confident. Our hooded cardigan tells the world that you are sporty yet elegant, with a flair for fashion. As you pose for a photo, you smile and feel like the whole world is smiling back at you. You're a star!
Sizes: XS 4, S 6-8, M 10-12, L 14-16, XL 18-20
Available colours:
Mountain Lilac
Coral Dust
Tarragon
Saltwater Blue
Oatmeal
Black
Done
My Computer

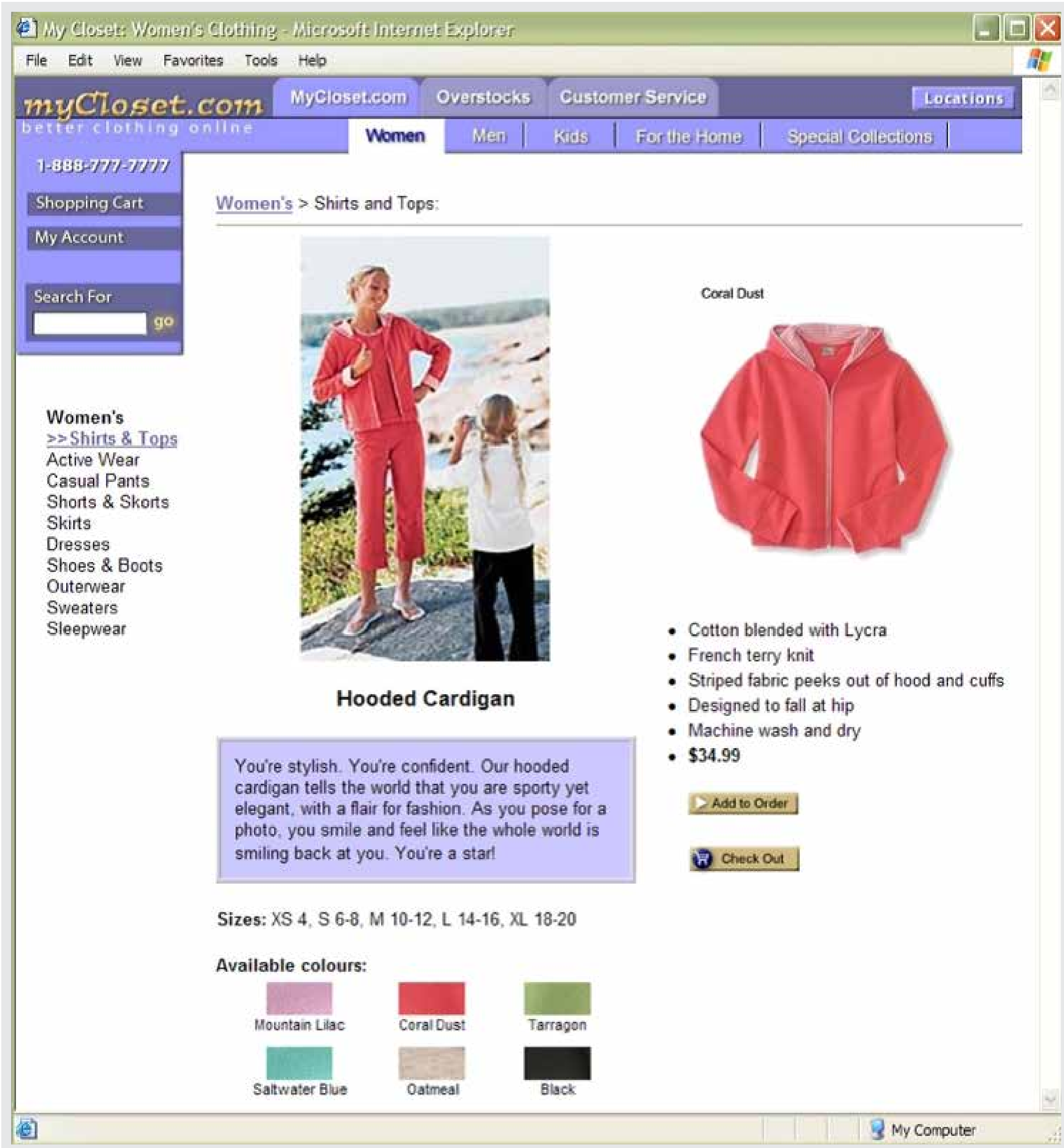


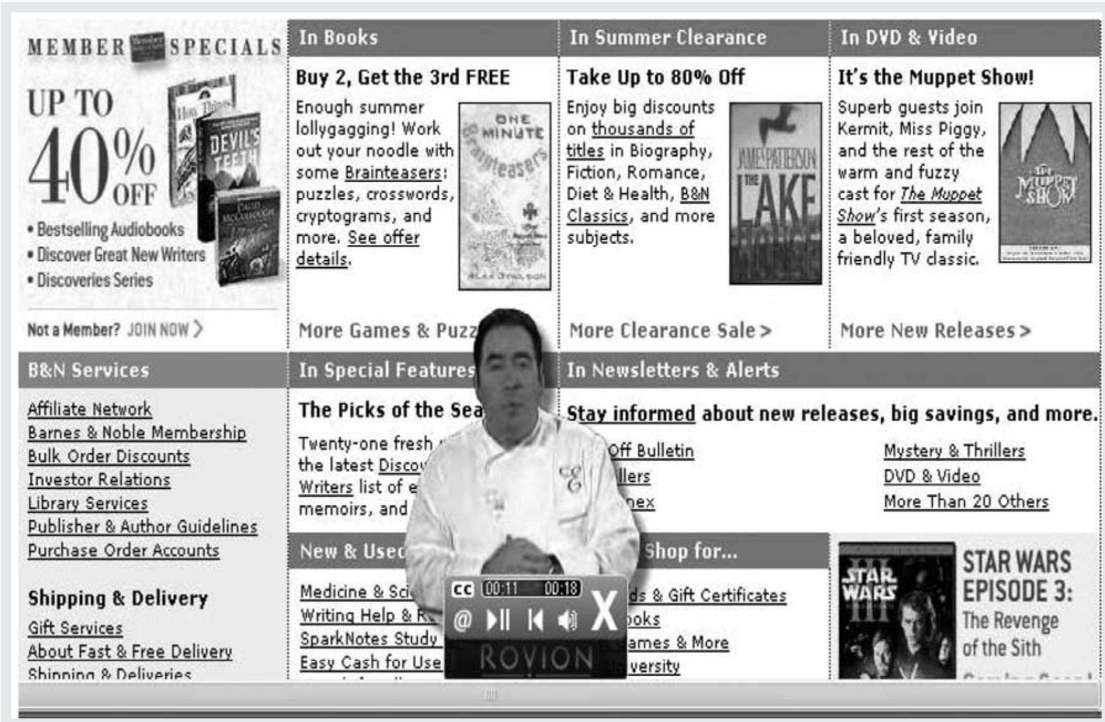



Figure 40.22 A-B-C: Difference Levels of Social Presence on Websites (in Hassanein and Head, 2007).

Hassanein and Head (2007) concluded that for shopping websites featuring apparel, higher levels of social presence, created in part through human figures, positively impacted perceived usefulness, trust, and enjoyment. Further, as the degree of perceived social presence increases, there is an increased impact on emotions and behavior (Argo et al., 2005). Alternately, the type of website determines whether or not the development of social presence is necessary. For instance, Hassanein and Head (2005-6) conducted another study in which an apparel website was compared to a website selling headphones. For the website selling headphones higher levels of social presence did not positively impact user attitudes, since the user was primarily seeking detailed product information.

### 40.4.2 Additional Forms of Online Social Influence

In related research, perceived "website socialness" was tested for relationships to perceived ease of use, perceived usefulness, enjoyment, and ultimately intention to use the website (Wakefield et al., 2011). To create greater variability in the experiment two types of websites were used: window blinds (utilitarian) and entertainers (hedonic). Participants were exposed to the website with either an interactive streaming video guide or the identical website but without the guide. The guide (Figure 40.23) featured four aspects of social cues: social role (the guide), human voice, language, and interactivity. The videotaped human guide that appeared on the screen invited users to interact with the guide, and was meant to make the pages "come alive". Results of this study supported all hypothesized relationships. Although contrary to the findings of Hassanein and Head (2005-6) there were no differences in how the model operated between the utilitarian and hedonic website conditions.




*Copyright © Barnes and Noble, and Rovion. All Rights Reserved. Used without permission under the Fair Use Doctrine (as permission could not be obtained). See the "Exceptions" section (and subsection "allRightsReserved-UsedWithoutPermission") on the page copyright notice.*

FIGURE 40.23: An Interactive Shopping Guide (in Wakefield et al., 2011).

Further, online shopping assistants are interactive technological artifacts that provide information and respond to online consumers. As "social actors" on the electronic stage, online shopping assistants potentially build a relationship with the user that results in enjoyment, trust, perceived usefulness, and ultimately in reuse intentions (Al-natour et al., 2005). If the personal shopping assistant is perceived by the user to be similar both in personality and behaviors, then it was more positively evaluated by the users. Additional studies by Benbasat and his colleagues examined "live help" functions through instant messaging or text chatting to facilitate interactions between users and online customer service representatives (Qui and Benbasat, 2005). Rather than text-based communication, users were exposed to socially rich environments which included computer-generated voice and humanoid avatars. These researchers found that text-to-speech voice delivered aloud through a 3-dimensional avatar significantly increased cognitive and emotional trust toward the customer service representative. Although these findings are several years old, the use of rich and personalized service agents still only appears selectively on websites. Figure 40.24 illustrates treatment websites featuring web pages with and without the Live Help interface.

Figure 40.24: A Live Help Interface with both Voice and Avatar (in Qui and Benbasat, 2005).

E-health is an emerging and important area for the application of social elements for website design. In a study of disclosure and personal health records, Lee and LaRose (2011) studied the impact of personalized social cues for two types of information disclosure (e.g. embarrassing information and descriptive information). Information disclosure was further tested related to user outcomes as positive (e.g. social trust and customization of their requests) or negative (embarrassment and information abuse). To manipulate the level of personalized social cues, in the high immediacy condition users were personally greeted using their first name (e.g. "Welcome Lissa") (Figure 40.25). In addition, the high immediacy website featured an interactive review session during which the user is praised and encouraged (e.g. "Excellent, Lissa! That is the right answer! Or "Try again, Lissa. You can do it!" Ibid. p. 339). Alternately, the low immediacy website did not host these social cues. An interesting finding from this study is that regardless of the type of information (embarrassing or descriptive), exposure to a high immediacy website with personalized social cues increased the level of information disclosure, signaling this as an important element of design in an e-health context.

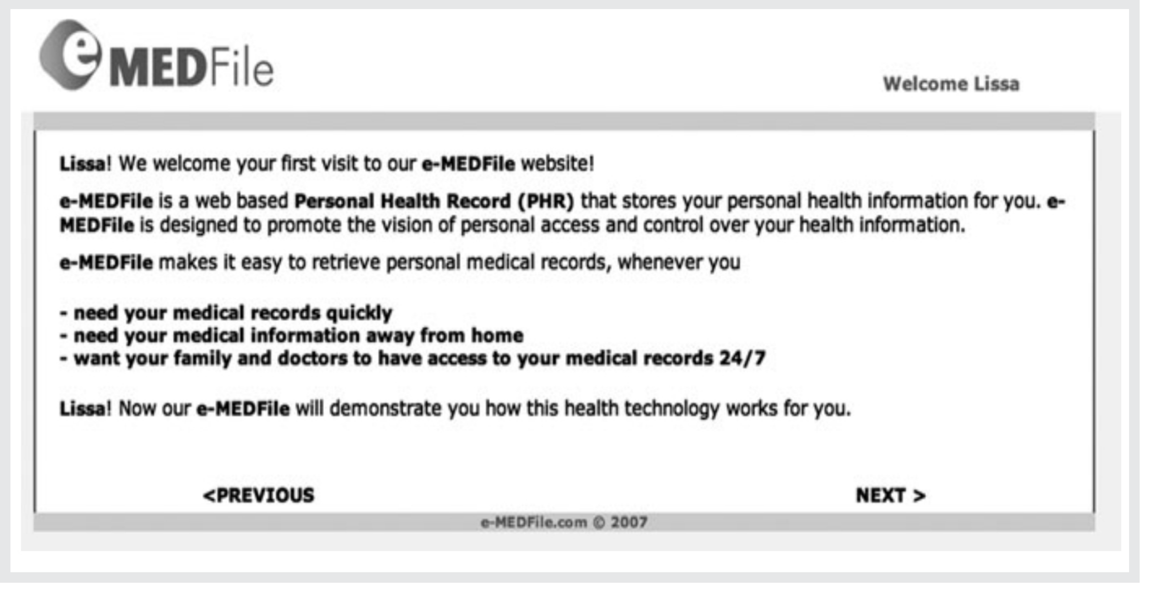





Figure 40.25: Screenshot for an Online Medical Website (in Lee and LaRose, 2011).

### 40.4.3 Men, Women and Social Cues

In 2013, "across all ages women interact more online than men".[20] Overall, men tend to have more positive evaluations of websites than women (Rodgers and Harris, 2003), including favorable attitudes toward online advertising (Parsa et al., 2011). In 2004, women were found to be less satisfied with websites, and trust them less than men (Garbarino and Strahilevitz, 2004), an outcome that remains current today.

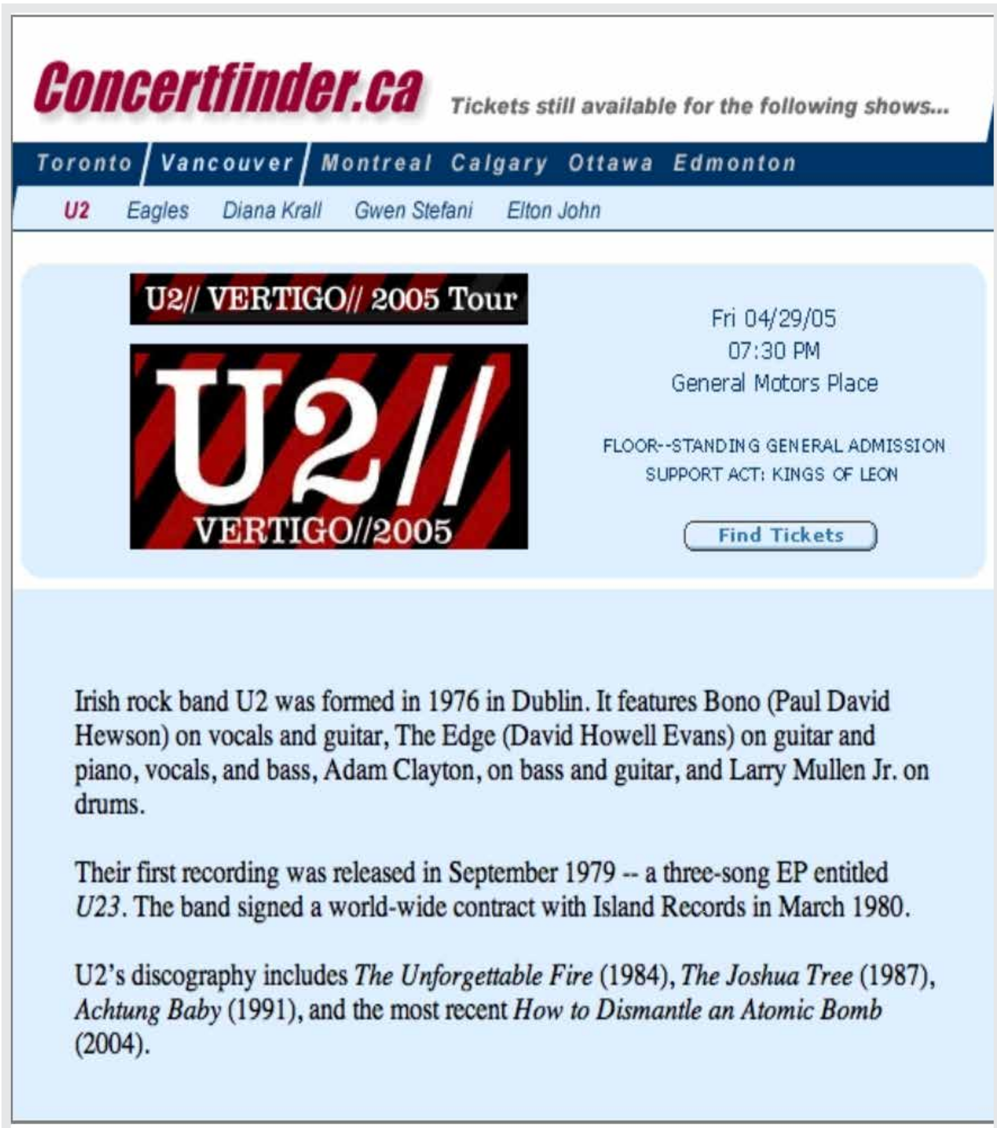
Generally, women have lower perceptions of websites than men regarding information richness, communication effectiveness, and the communication interface (Cyr and Bonanni, 2005). Women also have more negative evaluations of the presentation of product information and site organization, and are significantly less satisfied with navigation formats than men (Cyr and Bonanni, 2005).

20. Accessed 16 March 2013 at http://socialdriver.com/2013/01/the-tech-gender-gap/

Visualization characteristics such as preferences for color, shapes, and use of expert language vary between males and females (Ozdemir and Kilic, 2011). Further, differences between men and women were found for use of voice, color and language (Mahzari and Ahmadzadeh, 2013).

Men tend to focus on utilitarian aspects of shopping websites. Accurate product descriptions and fair pricing are very important for men (Chen and Hu, 2012;Ulbrich et al., 2011). Men like websites that allowed them to custom design products (Burke, 2002). Alternately, women dislike copious amounts of interactivity on the site, and find animations less meaningful than men. Women prefer websites with less clutter and fewer graphics. Women shoppers indicate they "must have" online product prices, a list of sale items, and personalized product discounts. More than men, women are receptive to electronic shopping lists, and having the website save a list of past purchases. Female web shoppers consider return labels (i.e. ability to return an item) more important than men (Ulbrich et al., 2011).

Concerning different social needs between men and women, the impact of social elements on websites is expected to be perceived differently between the sexes. To test this assumption, in one study five different levels of social presence were created for comparison between men and women (Cyr et al., 2007). All conditions featured the same content (for the same five performers or bands), but differed only in terms of social presence elements. Condition 1 was the basic treatment that included text and the band logo. In Condition 2, a photo of the band was included. Conditions 3 and 4 featured different interactive elements that allowed for discussions and reviews/ratings. Specifically, Condition 3 offered users the opportunity to open up a blank window and send a live chat message to other users assumed to be concurrently browsing the web page. The number of users browsing the current page was represented by a "presence indicator" consisting of a static image of several 'smiley face' icons. Condition 4 offered users the opportunity to see reviews from other users and write their own review for the performer/band. While Condition 3 was meant to simulate synchronous interaction, Condition 4 simulated asynchronous interaction with other website users. Finally, Condition 5 included all of the above mentioned features (text, logo, photo, synchronous chat, asynchronous reviews). (Figure 40.26(a-e)).

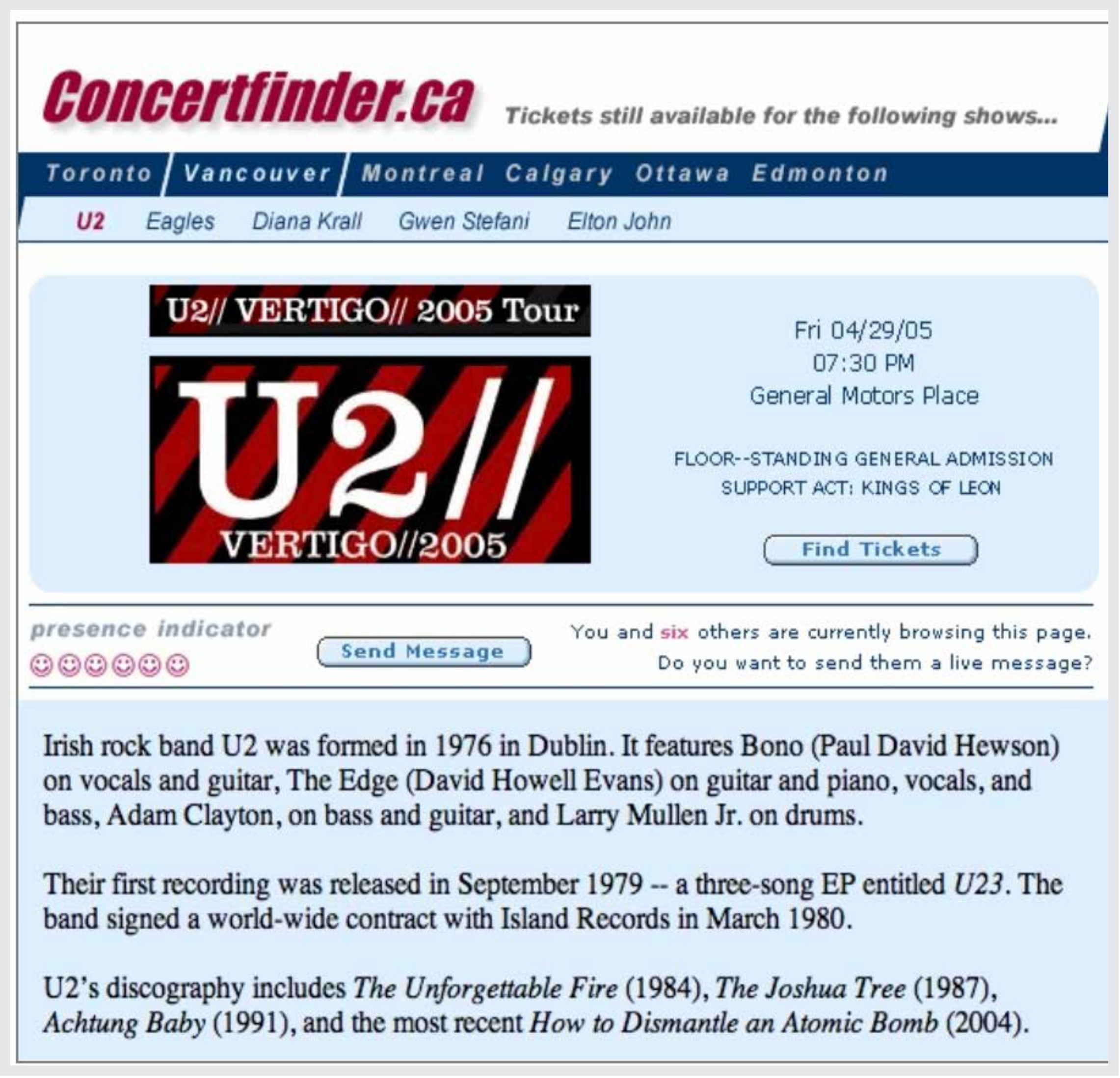





*Condition 1 (Basic).*

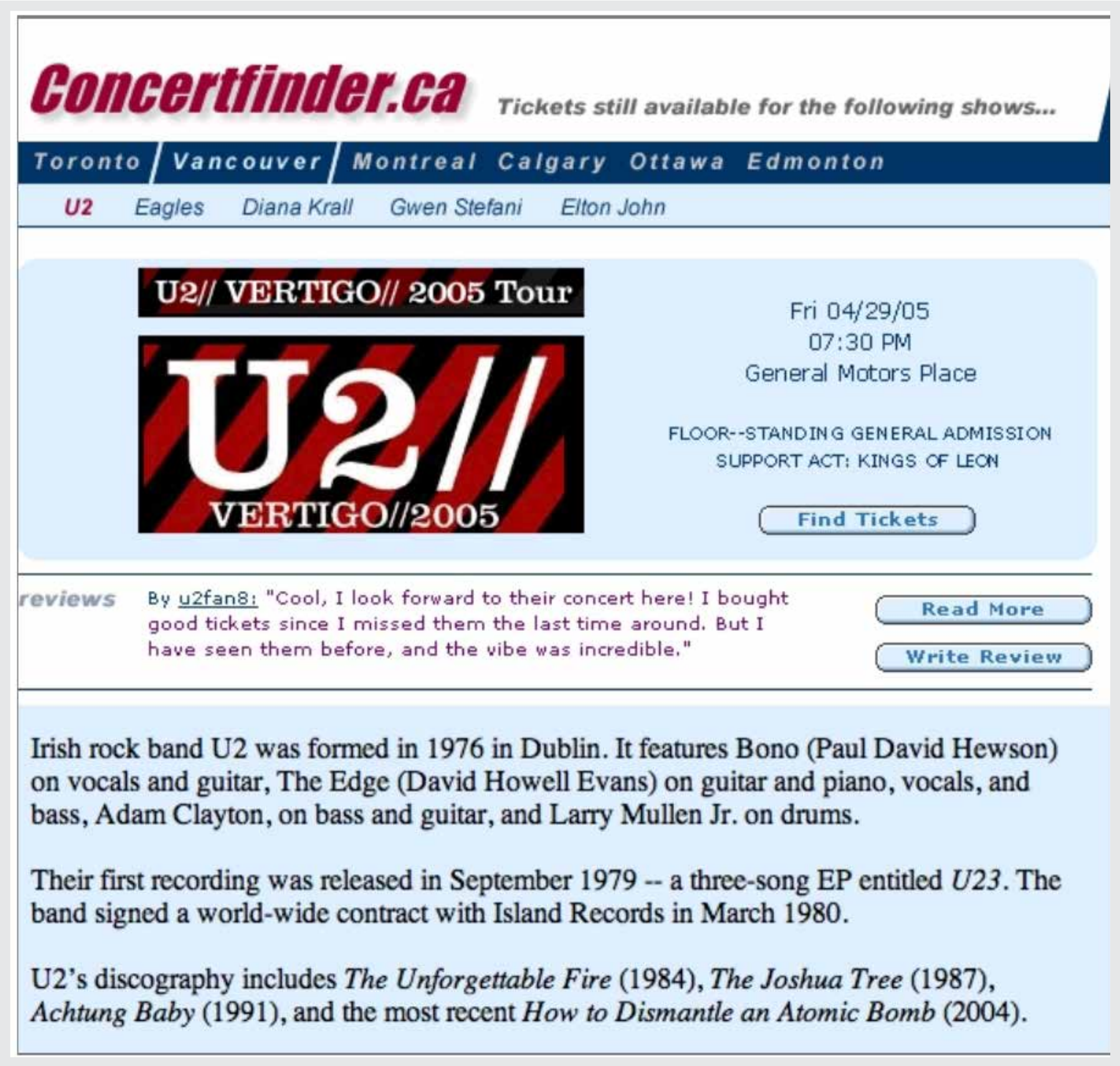



*Condition 2 (Photo).*






*Condition 3 (Synchronous chat).*






*Condition 4 (Asynchronous reviews).*




*Condition 5 (All).*

Figure 40.26: Difference Levels of Social Presence on Websites (in Cyr et al., 2007).

Results from this research indicated women sought hedonic content to "engage" them, whereas men focused more on information they felt was missing (Cyr et al., 2007). For a website with minimal social presence, women commented it was "boring, not enough pictures, no sense of vibrancy", that it had "no emotion, it does not evoke any response…a cold non-interactive site", that "visually, it is not very appealing at all. . . there should be more pictures", and that it was "not friendly". In contrast, when commenting on the higher social presence conditions, females noted they "felt relaxed and enjoyed reading it", "it aroused [their] curiosity of each band, making [them] want to listen to all of their music", and it felt "more like a party chat room than a cold, impersonal website just selling stuff/tickets" (ibid). In this study, perceived social presence was significantly related to perceived usefulness, trust, and ultimately loyalty for both men and women, although only for women did it result in a significant relationship of enjoyment to loyalty (ibid.).

Whether a man or a woman designs the website impacts user perceptions, although in 2013 only 5% to 20% of software developers are women[21]. Men have a preference for website pages produced by men, while women prefer web pages produced by women. For 30 male-produced and 30 female-produced websites, significant differences were found between the two sets of websites on 13 of 23 factors with respect to navigation and visual content. Websites designed by women had links to a larger number of topics than those designed by men. Language was used differently, with more references to competitiveness on male-produced websites. Of five language elements examined, there were significant differences on four of these elements with women more likely to use abbreviations and informal language. In particular, visual design varied between websites aimed at men versus women, with images of one's own gender appearing on the website. Women were more likely than males to use rounded rather than straight shapes, more colors, a horizontal layout, and informal images (Moss et al., 2006).

## 40.4.4 Social Elements of Website Design and Culture

Only a few studies have been conducted in which hedonic website design features are systematically modeled across diverse cultures, although recently research in

21. Accessed 16 March 2013 at http://socialdriver.com/2013/01/the-tech-gender-gap/

this area is beginning to emerge.Hassanein et al. (2009) aimed to determine if perceived social presence is culture specific or universal in online shopping settings. These researchers found that for Chinese and Canadian users, social presence led to perceptions of usefulness and enjoyment, but for Canadians social presence resulted in trust only.

As discussed earlier, Cyr and her colleagues conducted two separate research investigations with Canadian, German, and Japanese users regarding their reaction to visual design website elements. In the first study, survey data indicated that human images universally result in image appeal and perceived social presence; while interviews and eye-tracking data suggested participants from different cultures experience design images differently (Cyr et al., 2009). In the second study, website color appeal was found to be a significant determinant of website trust and satisfaction, with differences across cultures (Cyr et al., 2010.

### 40.4.5 Summary

Creating websites that are warm and inviting is clearly important to attract users and to encourage them to return or to purchase in the future. Despite this, many websites are lacking in this critical element. Social elements seem to be of particular importance to women, although they are also important to men. It is expected that if women dislike the online tone and format, then they will likely be less emotionally involved in the website or shopping experience. Therefore, there is merit in focusing on designing for a female audience in order to increase enjoyment, involvement, trust, and satisfaction for this group. One way to facilitate this process is to utilize more women in the design process. In addition to considering social elements of website design in domestic markets, there is room for further research and exploration of this topic in international contexts as well. As such, this presents a fruitful area for future studies.

## 40.5 FUTURE DIRECTIONS

While generally under-researched, the use of emotion in website design is gaining prominence and thus deserves further investigation. Identifying and systematically clarifying website design elements that result in positive emotional responses in the user has not only theoretical merit, but practical purpose as well. Taking a cue from the research in this paper, a number of avenues for future investigation or consideration are suggested:

- *Definition and Model Development of Emotion in Websites* — As noted early on, there are numerous definitions which all are related to emotion on websites. It is not entirely clear which website design elements result in which user emotions, although research in this area is evolving. Some researchers are proposing the development of integrated models for understanding affective responses (e.g. Lowry et al., 2013; Zhang, 2013) and work in this area will likely continue with useful outcomes for both academics and those responsible for the design of effective websites.
- *Instrument Validation* — Although there are numerous scales of usually four or five items that measure types of affective or emotional responses, there are few comprehensive instruments that serve this purpose. Hence, there is value in systematically developing and validating such an instrument.
- *Methodological Diversity* — Although research on website design and information design has typically relied on surveys with single or multiple item scales, more recently researchers have been diversifying their techniques to obtain different types of data. This might include research into emotion and website design that also includes neurophysiological techniques such as fMRI, an eye-tracking device

to determine exactly where and how users respond to emotional content on web pages, or a combination of these methodologies.

- *New Contexts for Online Emotion* — While there has been considerable attention devoted to utilitarian outcomes of website design in numerous contexts, there is a need to extend emotion in website design in a similar way. Areas for future research include emotion and website design in: e-government; e-health; mobile applications; gaming; social networking; and e-commerce, among others.
- *Emphasis on Diversity* – Generally, research has not carefully examined the emotional responses that different types of users have to website design elements. Based on the work already elaborated, it is important to factor age, gender, and culture into how users perceive websites. Enhanced knowledge in these areas will be of great importance to web designers and will ultimately have significant commercial value.

## 40.6 ACKNOWLEDGEMENT

Support for this project was provided by the Social Sciences and Humanities Research Council of Canada

a00ad

## 40.7 REFERENCES

*Agarwal, Ritu and Karahanna, Elena (2000): Time flies when you're having fun: Cognitive absorption and beliefs about information technology usage. In MIS Quarterly, 24 (4) pp. 665-694*

*Agarwal, Ritu and Venkatesh, Viswanath (2002): Assessing a firm's web presence: a heuristic evaluation procedure for the measurement of usability. In Information Systems Research, 13 (2) pp. 168-186*

*Al-Natour, Sameh, Benbasat, Izak and Cenfetelli, Ronald T. (2005): The role of similarity in e-commerce interactions: The case of online shopping assistants. In: Proceedings of the 4th Annual Workshop on HCI Research in MIS December 10, 2005, Las Vegas, USA. pp. 70-74*

*Anderson, Rolph E. and Srinivasan, Srini S. (2003): E-satisfaction and e-loyalty: a contingency framework. In Psychology & Marketing, 20 (2) pp. 123-138*

*Argo, Jennifer J., Dahl, Darren W. and Manchanda, Rajesh W. (2005): The influence of a mere social presence in a retail context. In Journal of Consumer Research, 32 (2) pp. 207-212*

*Bagozzi, Richard P. and Yi, Youjae (1989): On the use of structural equation models in experimental designs. In Journal of Marketing Research, 26 (3) pp. 271-284*

*Barber, Wendy and Badre, Albert (1998): Culturability: the merging of culture and usability. In: Proceedings of the 4th Conference on Human Factors and the Web June 5, 1998, Basking Ridge, USA.*

*Beaudry, Anne and Pinsonneault, Alain (2010): The Other Side of Acceptance: Studying the Direct and Indirect Effects of Emotions on Information Technology Use. In Management Information Systems Quarterly, 34 (4) pp. 689-710*

*Becker, Shirley A. (2002): An exploratory study on web usability and the internationalization of US e-business. In Journal of Electronic Commerce Research, 3 (4) pp. 265-278*

*Belanger, France, Hiller, Janine S. and Smith, Wanda J. (2002): Trustworthiness in electronic commerce: the role of privacy, security, and site attributes. In The Journal of Strategic Information Systems, 11 (3) pp. 245-270*

*Benbasat, Izak and Wang, Weiquan (2005): Trust In and Adoption of Online Recommendation Agents. In Journal of the Association for Information Systems, 6 (3) pp. 72-101*

*Berlin, Brent and Kay, Paul (1969): Basic Color Terms: Their Universality and Evolution. Berkeley, USA, University of California Press*

*Bhattacherjee, Anol (2002): Individual trust in online firms: scale development and initial test. In Journal of Management Information System, 19 (1) pp. 211-241*

*Blythe, Mark and Wright, Peter (2003): Introduction: From usability to enjoyment. In: Blythe, Mark, Overbeeke, Kees, Monk, Andrew F. and Wright, Peter (eds.). "Funology: From Usability to Enjoyment (XIII-XIX)". Dordrecht, Netherlands: Kluwer Academic Publishers*

*Bonnardel, Nathalie, Piolat, Annie and Bigot, Ludovic Le (2011): The impact of colour on Website appeal and users' cognitive processes. In Displays, 32 (2) pp. 69-80*

*Branthwaite, Alan (2002): Investigating the Power of Imagery in Mar- keting Communication: Evidence-Based Techniques. In Qualitative Market Research: An International Journal, 5 (3) pp. 164-171*

*Burke, Raymond (2002): Technology and the customer interface: what consumers want in the physical and virtual store. In Journal of the Academy of Marketing Science, 30 (4) pp. 411-432*

*Buscher, Georg, Cutrell, Edward and Morris, Meredith R. (2009): What do you see when you're surfing Using eye tracking to predict salient regions of web pages. In: Proceedings of the 27th International Conference Extended Abstracts on Human Factors in Computing Systems April 4-9, 2009, Boston, USA. pp. 21-30*

*Butcher, Ken, Sparks, Beverley and O'Callaghan, Frances (2001): Evaluative and relational influences on service loyalty. In International Journal of Service Industry Management, 12 (4) pp. 310-327*

*Carte, Traci A. and Russell, Craig J. (2003): In pursuit of moderation: nine common errors and their solutions. In MIS Quarterly, 27 (3) pp. 479-501*

*Cenfetelli, Ronald (2004): Getting in touch with our feelings towards technology. In: Academy of Management Best Paper Proceedings August, 2004, New Orleans, USA.*

*Chen, Sandy and Dhillon, Gurpreet (2003): Interpreting dimensions of consumer trust in e-commerce. In Information Technology and Management, 4 (2) pp. 303-318*

*Chen, I-Ching and Hu, Shueh-Cheng (2012): Gender differences in shoppers' behavioural reactions to ultra-low price tags at online merchants. In Electronic Commerce Research, 12 (4) pp. 485-504*

*Cheung, Christy M. K. and Lee, Matthew K. O. (2006): Understanding consumer trust in Internet shopping: A multidisciplinary approach. In JASIST - Journal of the American Society for Information Science and Technology, 57 (4) pp. 479-492*

*Childers, Terry L., Carr, Christopher L., Peck, Joann and Carson, Stephen (2001): Hedonic and utilitarian motivations for online retail shopping behavior. In Journal of Retailing, 77 (4) pp. 511-535*

*Corritore, Cynthia L., Kracher, Beverly and Wiedenbeck, Susan (2003): On-line trust: concepts, evolving themes, a model. In International Journal of Human-Computer Studies, 58 (6) pp. 737-758*

*Cronbach, Lee J. (1971): Test validation. In: Thorndike, Robert L. (ed.). "Educational Measurement". pp. 443-507*

*Csikszentmihalyi, Mihaly (1989): The Psychology of Optimal Experience. In Journal of Personality and Social Psychology, 56 (5) pp. 815-822*

*Cyr, Dianne (2008): Modelling Website Design across Cultures: Relationships to Trust, Satisfaction and E-loyalty. In Journal of Management Information Systems, 24 (4) pp. 47-72*

*Cyr, Dianne and Bonanni, Carole (2005): Gender and website design in e-business. In International Journal of Electronic Business, 3 (6) pp. 565-582*

*Cyr, Dianne and Trevor-Smith, Haizley (2004): Localization of Web Design: A Comparison of German, Japanese, and U.S. Website Characteristics. In JASIST-Journal of the American Society for Information Science and Technology, 55 (13) pp. 1-10*

*Cyr, Dianne, Head, Milena, Larios, Hector and Pan, Bing (2009b): Exploring human images in website design: a multi-method approach. In MIS Quarterly, 33 (3) pp. 539-566*

*Cyr, Dianne, Head, Milena and Larios, Hector (2010): Colour appeal in website design within and across cultures: A multi-method evaluation. In International Journal of Human-Computer Studies, 68 (1) pp. 1-21*

*Cyr, Dianne, Hassanein, Khaled, Head, Milena and Ivanov, Alex (2007): The role of social presence in establishing loyalty in e-Service environments. In Interacting with Computers, 19 (1) pp. 43-56*

*Cyr, Dianne, Bonanni, Carole, Bowes, John and Ilsever, Joe (2005): Beyond Trust: Website Design Preferences Across Cultures. In Journal of Global Information Management, 13 (4) pp. 25-54*

*Cyr, Dianne, Head, Milena M. and Ivanov, Alex (2006): Design aesthetics leading to m-loyalty in mobile commerce. In Information & Management, 43 (8) pp. 950-963*

*Cyr, Dianne, Head, Milena and Ivanov, Alex (2009a): Perceived interactivity leading to e-loyalty: Development of a model for cognitive-affective user responses. In International Journal of Human-Computer Studies, 67 (10) pp. 850-869*

*Dabholkar, Pratibha A. (1994): Incorporating Choice into an Attitudinal Framework: Analyzing Models of Mental Comparison Processes. In Journal of Customer Research, 21 (1) pp. 100-118*

*Dabholkar, Pratibha A. and Bagozzi, Richard P. (2002): An Attitudinal Model of Technology-Based Self-Service: Moderating Effects of Customer Traits and Situational Factors. In Journal of Academy of Marketing Science, 30 (3) pp. 187-201*

*Dellaert, Benedict G.C. and Dabholkar, Pratibha (2009): Increasing the Attractiveness of Mass Customization: The Role of Complementary Online Services and Range of Options. In International Journal of Electronic Commerce, 13 (2) pp. 43-70*

*Deng, Liqiong and Poole, Marshall S. (2010): Affect in Web interfaces: A study of the impacts of Web page visual complexity and order. In MIS Quarterly, 34 (4) pp. 711-730*

*Dholakia, Ruby R., Zhao, Miao, Dholakia, Nikhilesh and Fortin, David R. (2001): Interactivity and revisits to websites: a theoretical framework. In: Proceeding of 2001 AMA Winter Conference February 12, 2001, Scottsdale, USA. pp. 108-115*

*Dickinger, Astrid, Arami, Mitra and Meyer, David (2008): The role of perceived enjoyment and social norm in the adoption of technology with network externalities. In European Journal of Information Systems, 17 (1) pp. 4-11*

*Djamasbi, Soussan, Siegel, Marisa, Tullis, Tom and Dai, Rui (2010): Efficiency, Trust, and Visual Appeal: Usability Testing through Eye Tracking. In: HICSS 10 Proceedings of the 2010 43rd Hawaii International Conference on System Sciences January 5-8, 2010, Hawai,USA. pp. 1-10*

*Djamasbi, Soussan, Siegel, Marisa and Tullis, Tom (2012): Faces and Viewing Behavior: An Exploratory Investigation. In AIS Transactions on Human-Computer Interaction, 4 (3) pp. 190-211*

*Djamasbi, Soussan, Siegel, Marisa and Tullis, Tom (2011): Visual Hierarchy and Viewing Behavior: An Eye Tracking Study. In: Jacko, Julie A. (ed.). "Human-Computer Interaction: Design and Development Approaches". Berlin, Germany: Springerpp. 331-340*

*Djamasbi, Soussan, Siegel, Marisa and Tullis, Tom (2010a): Generation Y, web design, and eye tracking. In International Journal of Human-Computer Studies, 68 (5) pp. 307-323*

*Donath, Judith (2001): Mediated Faces. In: Proceedings of the 4th International Conference on Cognitive Technology Instruments of Mind August 6-9, 2001, Coventry, United Kingdom. pp. 373-390*

*Doohwang, Lee and LaRose, Robert (2011): The impact of personalized social cues of immediacy on consumers' information disclosure: a social cognitive approach. In Cyberpsychology, Behavior, and Social Networking, 14 (6) pp. 337-343*

*Eroglu, Sevgin A., Machleit, Karen A. and Davis, Lenita M. (2003): Empirical testing of a model of online store atmospherics and shopper responses. In Psychology and Marketing, 20 (2) pp. 139-150*

*Everard, Andrea and Galletta, Dennis F. (2006): How Presentation Flaws Affect Perceived Site Quality, Trust, and Intention to Purchase from an Online Store. In Journal of Management Information Systems, 22 (3) pp. 56-95*

*Falk, R. Frank and Miller, Nancy B. (1992): A primer for soft modeling. Akron, USA, The University of Akron Press*

*Fiore, Ann M., Jin, Hyun-Jeong and Kim, Jihyun (2005): For fun and profit: Hedonic value from image interactivity and responses toward an online store. In Psychology and Marketing, 22 (8) pp. 669-694*

*Flavián, Carlos, Guinalíu, Miguel and Gurrea, Raquel (2006): The role played by perceived usability, satisfaction and consumer trust on website loyalty. In Information & Management, 43 (1) pp. 1-14*

*Fogg, B.J., Soohoo, Cathy and Danielsen, David (2002). How Do People Evaluate a Web Site's Credibility? Results from a Larger Study. Stanford University http://simson.net/ref/2002/stanfordPTL.pdf*

*Fornell, Claes (1982): A Second Generation of Multivariate Analysis: Methods. New York, USA, Praeger*

*Fortin, David R. and Dholakia, Ruby R. (2003): Interactivity and vividness effects on social presence and involvement with a web-based advertisement. In Journal of Business Research, 58 (3) pp. 387-396*

*Garbarino, Ellen and Strahilevitz, Michal (2004): Gender differences in the perceived risk of buying online and the effects of receiving a site recommendation. In Journal of Business Research, 57 (7) pp. 768-775*

*Garrett, Jesse J. (2003): The Elements of User Experience: User-Centered Design for the Web. Thousand Oaks, USA, New Riders*

*Gefen, David (2000): E-commerce: the role of familiarity and trust. In OMEGA, 28 (6) pp. 725-737*

*Gefen, David and Straub, Detmar W. (2003): Managing User Trust in B2C E-Services. In e-Service Journal, 2 (2) pp. 7-24*

*Gefen, David and Straub, Detmar W. (2004): Consumer trust in B2C e-Commerce and the importance of social presence: experiments in e-Products and e-Services. In OMEGA, 32 (6) pp. 407-424*

*Gefen, David, Karahanna, Elena and Straub, Detmar W. (2003): Trust and TAM in Online Shopping: An Integrated Model. In MIS Quarterly, 27 (1) pp. 51-90*

*Giddens, Anthony (1990): The Consequences of Modernity. Stanford, USA, Stanford University Press*

*Goldberg, Joseph H. and Kotval, Xerxes P. (1999): Computer interface evaluation using eye movements: methods and constructs. In International Journal of Industrial Ergonomics, 24 (6) pp. 631-645*

*Goldberg, Joseph H., Stimson, Mark J., Lewenstein, Marion, Scott, Neil and Wichansky, Anna M. (2002): Eye tracking in web search tasks: design implications. In: Duchowski, Andrew T., Vertegaal, Roel and Senders, John W. (eds.) ETRA 2002 - Proceedings of the Eye Tracking Research and Application Symposium March 25-27, 2002, New Orleans, Louisiana, USA. pp. 51-58*

*Gorn, Gerald J., Chattopadhyay, Amitava, Yi, Tracey and Dahl, Darren W. (1997): Effects of colour as an executional cue in advertising: they're in the shade. In Management Science, 43 (10) pp. 1387-1400*

*Gretzel, Ulrike, Fesenmaier, Daniel R., Formica, Sandro and O'Leary, Joseph T. (2006): Searching for the Future: Challenges Faced by Destination Marketing Organizations. In Journal of Travel Research, 45 (2) pp. 116-126*

*Griffith, David A., Krampf, Robert F. and Palmer, Jonathan W. (2001): The Role of Interface in Electronic Commerce: Customer Involvement with Print Versus Online Catalogs. In International Journal of Electronic Commerce, 5 (4) pp. 135-153*

*Ha, Imsook, Yoon, Youngseog and Choi, Munkee (2007): Determinants of adoption of mobile games under mobile broadband wireless access environment. In Information and Management, 44 (3) pp. 276-286*

*Hassanein, Khaled and Head, Milena (2006): The Impact of Infusing Social Presence in the Web Interface: An Investigation Across Product Types. In International Journal of Electronic Commerce, 10 (2) pp. 31-55*

*Hassanein, Khaled and Head, Milena (2007): Manipulating perceived social presence through the web interface and its impact on attitude towards online shopping. In International Journal of Human-Computer Studies, 65 (8) pp. 689-708*

*Hassanein, Khaled, Head, Milena and Ju, Chunhua (2009): A cross-cultural comparison of the impact of Social Presence on website trust, usefulness and enjoyment. In International Journal of Electronic Business, 7 (6) pp. 625-641*

Hassenzahl, Marc (2006): *Hedonic, Emotional, and Experiential Perspectives on Product Quality. In: Ghaoui, Claude (ed.). "Encyclopedia of human computer interaction". Idea Group Referencepp. 266-272*

Hassenzahl, Marc (2001): *The Effect of Perceived Hedonic Quality on Product Appealingness. In International Journal of Human-Computer Interaction, 13 (4) pp. 481-499*

Heijden, Hans van der (2004): *User Acceptance of Hedonic Information Systems. In MIS Quarterly, 28 (4) pp. 695-704*

Heijden, Hans van der (2003): *Factors influencing the usage of websites: the case of a generic portal in The Netherlands. In Information and Management, 40 (6) pp. 541-549*

Hoffman, Donna L. and Novak, Thomas P. (1996): *Marketing in hypermedia computer-mediated environments: Conceptual foundations. In Journal of Marketing, 60 (3) pp. 50-68*

Hofstede, Geert (1980): *Culture's consequences: International differences in work-related values. Beverly Hills, USA, Sage Publications*

Huang, Ming-Hui (2006): *Flow, enduring, and situational involvement in the Web environment: A tripartite second-order examination. In Psychology & Marketing, 23 (5) pp. 383-411*

Ivanov, Alex, Erickson, Thomas and Cyr, Dianne (2006): *Plot-polling: Collaborative Knowledge Visualization for Online Discussions. In: Tenth International Conference on Information Visualization July 5-6, 2006, London, United Kingdom. pp. 205-210*

Jahng, Namsook, Krug, Don and Zhang, Zuochen (2007): *Student Achievement in Online Distance Education Compared to Face-to-Face Education. In European Journal of Open, Distance and E-Learning,*

Jahng, Jungjoo J., Jain, Hemant K. and Ramamurthy, K. (2002): *Personality traits and effectiveness of presentation of product information in e-business systems. In European Journal of Information Systems, 11 (3) pp. 181-195*

Jarvenpaa, Sirkka L., Tractinsky, Noam and Vitale, Michael (2000): *Consumer trust in an Internet store. In Information Technology and Management, 1 (2) pp. 45-71*

Jarvenpaa, Sirkka L., Tractinsky, Noam and Saarinen, Lauri (1999): *Consumer Trust in an Internet Store: A Cross-Cultural Validation. In J. Computer-Mediated Communication, 5 (2)*

Jiang, Zhenhui and Benbasat, Izak (2007): *The Effects of Presentation Formats and Task Complexity on Online Consumers' Product Understanding. In MIS Quarterly, 31 (3) pp. 475-500*

*Jiang, Zhenhui, Chan, Jason, Tan, Bernard C.Y. and Chua, Wei Siong (2010): Effects of Interactivity on Website Involvement and Purchase Intention. In Journal of the Association for Information Systems, 11 (1) pp. 34-59*

*Johnson, Daniel M. and Wiles, Janet (2003): Effective Affective User Interface Design in Games. In Ergonomics, 46 (13) pp. 1332-1345*

*Johnson, Grace J., Bruner, Gordon C. and Kumar, Anand (2006): Interactivity and its facets revisited: Theory and empirical test. In Journal of Advertising, 35 (4) pp. 35-52*

*Kamp, Edward and MacInnis, Deborah (1995): Characteristics of portrayed emotions in commercials: when does what is shown in ads affect viewers?. In Journal of Advertising Research, 22 (4) pp. 19-28*

*Karvonen, Kristiina (2000): The Beauty of Simplicity. In: Proceedings of the 2000 ACM Conference on Universal Usability 2000. pp. 85-90*

*Kay, Paul, Berlin, Brent and Merrifield, William (1991): Biocultural implications of systems of colour naming. In Journal of Linguistic Anthropology, 1 (1) pp. 12-25*

*Kim, Dongmin and Benbasat, Izak (2006): The Effects of Trust-Assuring Arguments on Consumer Trust in Internet Stores: Application of Toulmin's Model of Argumentation. In Information Systems Research, 17 (3) pp. 286-300*

*Kim, Jinwoo and Moon, Jae Yun (1998): Designing Towards Emotional Usability in Customer Interfaces -- Trustworthiness of Cyber-Banking System Interfaces. In Interacting with Computers, 10 (1) pp. 1-29*

*Kim, Jinwoo, Lee, Jooeun and Choi, Dongseong (2003): Designing emotionally evocative homepages: an empirical study of the quantitative relations between design factors and emotional dimensions. In International Journal of Human-Computer Studies, 59 (6) pp. 899-940*

*Kim, Hee-Woong, Chan, Hock Chuan and Chan, Yee Pia (2007): A balanced thinking-feelings model of information systems continuance. In International Journal of Human-Computer Studies, 65 (6) pp. 511-525*

*Komiak, Sherrie Y. X. and Benbasat, Izak (2006): The Effects of Personalization and Familiarity on Trust and Adoption of Recommendation Agents. In MIS Quarterly, 30 (4) pp. 941-960*

*Komiak, Sherrie X. and Benbasat, Izak (2004): Understanding Customer Trust in Agent-Mediated Electronic Commerce, Web-Mediated Electronic Commerce, and Traditional Commerce. In Information Technology and Management, 5 (1) pp. 181-207*

Koufaris, Marios (2002): Applying the Technology Acceptance Model and Flow Theory to Online Consumer Behavior. In Information Systems Research, 13 (2) pp. 205-223

Koufaris, Marios and Hampton-Sosa, William (2004): The development of initial trust in an online company by new customers. In Information & Management, 41 (3) pp. 377-397

Koufaris, Marios, Kambil, Ajit and Labarbera, Priscilla Ann (2002): Customer Behavior in Web-Based Commerce: An Empirical Study. In International Journal of Electronic Commerce, 6 (2) pp. 115-138

Kulkarn, R.B, Rajeshwarkar, Pooja R. and Dixit, S.K (2012): Cultural Analysis of Indian Websites. In International Journal of Computer Applications, 38 (8) pp. 15-21

Kumar, Nanda and Benbasat, Izak (2002): Para-social Presence: A Re-conceptualization of 'Social Presence' to Capture the Relationship between a Web Site and Her Visitors. In: HICSS 2002 2002. p. 13

Lam, Renee and Lim, Kai (2004): Emotions in Online Shopping: Fulfilling Customer's Needs through Providing Emotional Features and Customizing Website Features. In: Proceedings of the International Conference on Information Systems, ICIS 2004 December 12-15, 2004, Washington, USA. pp. 877-888

Lam, Shun Yin, Shankar, Venkatesh, Erramilli, M. Krishna and Murthy, Bvsan (2004): Customer Value, Satisfaction, Loyalty, and Switching Costs: An Illustration From a Business-to-Business Service Context. In Journal of the Academy of Marketing Science, 32 (3) pp. 293-311

Lankton, Nancy K. and Wilson, E. Vance (2007): Factors Influencing Expectations of E-Health Services Within a Direct-Effects Model of User Satisfaction. In e-Service Journal, 5 (2) pp. 85-111

Latornia, M.J. and Happ, A.J. (1987): The effective use of colour for text on the IBM 5153 colour display: random access II. In: Proceedings of the Human Factors Society 31st Annual Meeting 1987. pp. 1091-1095

Lavie, Talia and Tractinsky, Noam (2004): Assessing dimensions of perceived visual aesthetics of web sites. In International Journal of Human-Computer Studies, 60 (3) pp. 269-298

Lee, ThaeMin (2005): The impact of perceptions of interactivity on customer trust and transaction intentions in mobile commerce. In Journal of Electronic Commerce Research, 6 (3) pp. 165-180

Lee, Jungwon, Kim, Jinwoo and Moon, Jae Yun (2000): What Makes Internet Users Visit Cyber Stores Again? Key Design Factors for Customer Loyalty. In: Turner, Thea, Szwillus, Gerd, Czerwinski, Mary, Peterno, Fabio and Pemberton, Steven (eds.) Proceedings of the ACM CHI 2000 Human Factors in Computing Systems Conference April 1-6, 2000, The Hague, The Netherlands. pp. 305-312

Li, Han, Sarathy, Rathindra and Zhang, Jie (2008): The Role of Emotions in Shaping Consumers' Privacy Beliefs About Unfamiliar Online Vendors. In Journal of Information Privacy and Security, 4 (3) pp. 36-62

Lichtlé, Marie-Christine (2007): The effect of an advertisement's colour on emotions evoked by an ad and attitude towards the ad. In International Journal of Advertising, 26 (1) pp. 37-62

Lim, Eric T.K. and Cyr, Dianne (2009): Modeling Hedonic Consumption Behaviors in Online Shopping. In: Proceedings for the Eighth Pre-ICIS HCI Research in MIS Workshop HCI/MIS'09 December 14, 2009, Phoenix, USA.

Lin, Chieh-Peng and Bhattacherjee, Anol (2010): Extending technology usage models to interactive hedonic technologies: a theoretical model and empirical test. In Information Systems Journal, 20 (2) pp. 163-181

Lin, Chieh-Peng and Bhattacherjee, Anol (2008): Elucidating Individual Intention to Use Interactive Information Technologies: The Role of Network Externalities. In International Journal of Electronic Commerce, 13 (1) pp. 85-108

Lin, Aleck, Gregor, Shirley and Ewing, Michael (2008): Developing a scale to measure the enjoyment of Web experiences. In Journal of Interactive Marketing, 22 (4) pp. 40-57

Lindgaard, Gitte and Dudek, Cathy (2003): What is this evasive beast we call user satisfaction?. In Interacting with Computers, 15 (3) pp. 429-452

Lindgaard, Gitte, Fernandes, Gary, Dudek, Cathy and Brown, J. (2006): Attention web designers: You have 50 milliseconds to make a good first impression. In Behaviour and Information Technology, 25 (2) pp. 115-126

Lisetti, Christine L. and Nasoz, Fatma (2002): MAUI: a multimodal affective user interface. In: ACM Multimedia 2002 2002. pp. 161-170

Liu, Chang, Marchewka, Jack T. and Ku, Catherina (2004): American and Taiwanese Perceptions Concerning Privacy, Trust, and Behavioral Intentions in Electronic Commerce. In Journal of Global Information Management, 12 (1) pp. 18-40

Liu, Chang, Arnett, Kirk P., Capella, Louis M. and Taylor, Ronald D. (2001): Key dimensions of Web Design Quality as Related to Consumer Responses. In Journal of Computer Information Systems, 42 (1) pp. 70-83

*Loiacono, Eleanor and Djamasbi, Soussan (2010): Moods and Their Relevance to Systems Usage Models within Organizations: An Extended Framework. In Transactions on Human-Computer Interaction, 2 (2) pp. 55-72*

*Loiacono, Eleanor, Watson, Richard and Goodhue, Dale (2007): WebQual: An Instrument for Consumer Evaluation of Web Sites. In International Journal of Electronic Commerce, 11 (3) pp. 51-87*

*Lombard, Matthew and Ditton, Theresa (1997): At the Heart of It All: The Concept of Presence. In J. Computer-Mediated Communication, 3 (2)*

*Lowry, Paul B., Gaskin, James, Twyman, Nathan, Hammer, Bryan and Roberts, Tom (2013): Proposing the Hedonic-Motivation System Adoption Model (HMSAM) to Increase Understanding of Adoption of Hedonically Motivated Systems. In Journal of the Association for Information Systems,*

*Luarn, Pin and Lin, Hsin-Hui (2003): A customer loyalty model for e-service context. In Journal of Electronic Commerce Research, 4 (4) pp. 156-167*

*Madden, Thomas J., Hewett, Kelly and Roth, Martin S. (2000): Managing images in different cultures: a cross-national study of colour meanings and preferences. In Journal of International Marketing, 8 (4) pp. 90-107*

*Mahzari, Anahita and Ahmadzadeh, Marzieh (2013): Finding Gender Preferences in E-Commerce Website Design by an Experimental Approach. In International Journal of Applied Information Systems, 5 (2) pp. 35-40*

*Marcus, Aaron and Gould, Emilie West (2000): Crosscurrents: cultural dimensions and global Web user-interface design. In Interactions, 7 (4) pp. 32-46*

*McKnight, D. Harrison, Choudhury, Vivek and Kacmar, Charles (2002): Developing and Validating Trust Measures for e-Commerce: An Integrative Typology. In Information Systems Research, 13 (3) pp. 334-359*

*McQuarrie, Edward F. and Phillips, Barbara J. (2005): Indirect Persuasion in Advertising: How Consumers Process Metaphors Presented in Pictures and Words. In Journal of Advertising, 34 (2) pp. 7-20*

*Meyers-Levy, Joan and Peracchio, Laura A. (1995): Understanding the Effects of Color: How the Correspondence between Available and Required Resources Affects Attitudes. In Journal of Consumer Research, 22 (2) pp. 121-138*

*Meyers-Levy, Joan and Peracchio, Laura A. (1992): Getting an Angle in Advertising: The Effect of Camera Angle on Product Evaluations. In Journal of Marketing Research, 29 (4) pp. 454-461*

*Moss, Gloria, Gunn, Rod and Heller, Jonathan (2006): Some men like it black, some women like it pink: consumer implications of differences in male and female website design. In Journal of Consumer Behavior, 5 (4) pp. 328-341*

*Nielsen, Jakob (2004): Designing for web usability. Indianapolis,USA, New Riders Publications*

*Nielsen, Jakob and Galdo, Elisa M. del (1996): International User Interfaces. New York, USA, Wiley Computer Publishing*

*Noiwan, Jantawan and Norcio, Anthony F. (2006): Cultural differences on attention and perceived usability: Investigating color combinations of animated graphics. In International Journal of Human-Computer Studies, 64 (2) pp. 103-122*

*Oh, Jungmi, Fiorito, Susan S., Cho, Hira and Hofacker, Charles F. (2008): Effects of design factors on store image and expectation of merchandise quality in web-based stores. In Journal of Retailing and Consumer Services, 15 (4) pp. 237-249*

*Ozdemir, Erkan and Kilic, Serkan (2011): Young Consumers' Perspectives of Website Visualization: A Gender Perspective. In Business and Economics Research Journal, 2 (2) pp. 41-60*

*Palmer, Jonathan W. (2002): Web Site Usability, Design, and Performance Metrics. In Information Systems Research, 13 (2) pp. 151-167*

*Park, Jihye, Lennon, Sharron J. and Stoel, Leslie (2005): On-line product presentation: Effects on mood, perceived risk, and purchase intention. In Psychology & Marketing, 22 (9) pp. 695-719*

*Parsa, Faramarz, Tomic, Igor, Angelidis, John and Ibrahim, Nabil A. (2011): Gender effects on e-commerce. In Journal of Academy of Business and Economics, 11 (4) pp. 219-223*

*Qiu, Lingyun and Benbasat, Izak (2005): Online Consumer Trust and Live Help Interfaces: The Effects of Text-to-Speech Voice and Three-Dimensional Avatars. In International Journal of Human-Computer Interaction, 19 (1) pp. 75-94*

*Rafaeli, Sheizaf and Noy, Avi (2005): Social Presence: Influence on Bidders in Internet Auctions. In Electronic Markets, 15 (2) pp. 158-175*

*Rattanawicha, Pimmanee and Esichaikul, Vatcharaporn (2005): What makes websites trustworthy? A two-phase empirical study. In International Journal of Electronic Commerce, 3 (2) pp. 110-136*

*Riedl, René, Hubert, Marco and Kenning, Peter (2010): Are There Neural Gender Differences in Online Trust? An fMRI Study on the Perceived Trustworthiness of eBay Offers. In MIS Quarterly, 34 (2) pp. 397-428*

*Riegelsberger, Jens (2002): The Effect of Facial Cues on Trust in E-Commerce Systems. In: Proceedings of the 16th British Human Computer Interaction Conference September 1-5, 2002, London, United Kingdom.*

*Riegelsberger, Jens, Sasse, Martina Angela and McCarthy, John D. (2003): Shiny happy people building trust?: photos on e-commerce websites and consumer trust. In: Cockton, Gilbert and Korhonen, Panu (eds.) Proceedings of the ACM CHI 2003 Human Factors in Computing Systems Conference April 5-10, 2003, Ft. Lauderdale, Florida, USA. pp. 121-128*

*Rivard, Suzanne and Huff, Sid L. (1988): Factors of Success for End-User Computing. In Communications of the ACM, 31 (5) pp. 552-561*

*Robins, David and Holmes, Jason (2008): Aesthetics and credibility in web site design. In Information Processing & Management, 44 (1) pp. 386-399*

*Rodgers, Shelly and Harris, Mary A. (2003): Gender and e-commerce: an exploratory study. In Journal of Advertising Research, 43 (3) pp. 322-329*

*Roseman, Ira J., Antoniou, Ann A. and Jose, Paul E. (1996): Appraisal Determinants of Emotions: Constructing a More Accurate and Comprehensive Theory. In Cognition and Emotion, 10 (3) pp. 241-277*

*Rosen, Deborah E. and Purinton, Elizabeth (2004): Website design: Viewing the web as a cognitive landscape. In Journal of Business Research, 57 (7) pp. 787-794*

*Russell, James A. (1980): A circumplex model of affect. In Journal of Personality and Social Psychology, 39 (6) pp. 1161-1178*

*Santosa, Paulus I., Wei, Kwok K. and Chan, Hock C. (2005): User involvement and user satisfaction with information-seeking activity. In European Journal of Information Systems, 19 (5) pp. 367-377*

*Schenkman, Bo N. and Jonsson, Fredrik U. (2000): Aesthetics and preferences of web pages. In Behaviour & Information Technology, 19 (5) pp. 367-377*

*Scherer, Klaus R. (1997): Profiles of Emotion-antecedent Appraisal: Testing Theoretical Predictions across Cultures. In Cognition and Emotion, 11 (2) pp. 113-150*

*Seidenspinner, Margarete and Theuner, Gabriele (2007): Intercultural aspects of online communication a comparison of mandarin-speaking, US, Egyptian and German user preferences. In Journal of Business Economics and Management, 8 (2) pp. 101-109*

Shen, Kathy N. and Khalifa, Mohamed (2009): Design for social presence in online communities: a multidimensional approach. In AIS Transactions on Human-Computer Interaction, 1 (2) pp. 33-54

Singh, Nitish, Zhao, Hongxin and Hu, Xiaorui (2003): Cultural Adaptation on the Web: A Study of American Companies' Domestic and Chinese Websites. In Journal of Global Information Management, 11 (3) pp. 63-80

Singh, Nitish, Toy, Daniel R. and Wright, Lauren K. (2009): A diagnostic framework for measuring Web-site localization. In Thunderbird International Business Review, 51 (3) pp. 281-295

Smith, Craig A. and Lazarus, Richard S. (1993): Appraisal components, core relational themes, and the emotions. In Cognition and Emotion, 7 (3) pp. 233-269

Steinbrück, Ulrike, Schaumburg, Heike, Duda, Sabrina and Krüger, Thomas (2002): A picture says more than a thousand words: photographs as trust builders in e-commerce websites. In: Terveen, Loren (ed.) Proceedings of CHI 2002 Conference on Human Factors in Computing Systems April 20-25, 2002, Minneapolis,USA. pp. 748-749

Straub, Detmar W. (1989): Validating Instruments in MIS Research. In MIS Quarterly, 12 (2) pp. 147-170

Sun, Huatong (2001): Building a culturally-competent corporate web site: an exploratory study of cultural markers in multilingual web design. In: IEEE ACM 19th International Conference on Computer Documentation 2001. pp. 95-102

Sun, Heshan and Zhang, Ping (2006): The role of moderating factors in user technology acceptance. In International Journal of Human-Computer Studies, 64 (2) pp. 53-78

Swinyard, William R. (1993): The Effects of Mood, Involvement, and Quality of Store Experience on Shopping Intention. In Journal of Consumer Research, 20 (2) pp. 271-280

Szymanski, David M. and Hise, Richard T. (2000): E-satisfaction: an initial examination. In Journal of Retailing, 76 (3) pp. 309-322

Tarasewich, Peter (2003): Wireless devices for mobile commerce: user interface design and usability. In: Mennecke, Brian E. and Strader, Troy J. (eds.). "Mobile commerce: technology, theory, and applications". Hershey, USA: Idea Group Publishingpp. 26-50

Tractinsky, Noam (2004): Toward the Study of Aesthetics in Information Technology. In: Proceedings from the 25th Annual International Conference on Information Systems December 9-12, 2004, Washington, USA. pp. 12-15

*Tractinsky, Noam (2006): Aesthetics in Information Technology: Motivation and Future Research Directions. In: Zhang, Ping and Galletta, Dennis (eds.). "Human-Computer Interaction and Management Information Systems: Foundations (Advances in Management Information Systems)". M.E. Sharpe*

*Tullis, Tom, Siegel, Marisa and Sun, Emily (2009): Are people drawn to faces on webpages?. In: Proceedings of ACM CHI 2009 Conference on Human Factors in Computing Systems 2009. pp. 4207-4212*

*Ulbrich, Frank, Christensen, Tina and Stankus, Linda (2011): Gender-specific on-line shopping preferences. In Electronic Commerce Research, 11 (2) pp. 181-199*

*Venkatesh, Viswanath (2000): Determinants of Perceived Ease of Use: Integrating Control, Intrinsic Motivation, and Emotion into the Technology Acceptance Model. In Information Systems Research, 11 (4) pp. 342-365*

*Vyncke, Femke and Brengman, Malaika (2010): Are Culturally Congruent Websites More Effective?An overview of a decade of empirical evidence. In Journal of Electronic Commerce Research, 11 (1) pp. 14-29*

*Wakefield, Robin L. and Whitten, Dwayne (2006): Mobile computing: a user study on hedonic/utilitarian mobile device usage. In European Journal of Information Systems, 15 (3) pp. 292-300*

*Wakefield, Robin L., Wakefield, Kirk L., Baker, Julie and Wang, Liz C. (2011): How website socialness leads to website use. In European Journal of Information Systems, 20 (1) pp. 118-249*

*Warner, Richard (1980): Enjoyment. In Philosophical Review, 89 (4) pp. 507-526*

*Witmer, Bob G., Jerome, Christian J. and Singer, Michael J. (2005): The Factor Structure of the Presence Questionnaire. In Presence: Teleoperators and Virtual Environments, 14 (3) pp. 298-312*

*Yoo, Youngjin and Alavi, Maryam (2001): Media and group cohesion: relative influences on social presence, task participation, and group consensus. In MIS Quarterly, 25 (3) pp. 371-390*

*Yoon, Sung-Joon (2002): The antecedents and consequences of trust in online-purchase decisions. In Journal of Interactive Marketing, 16 (2) pp. 47-63*

*Zhang, Ping (2013): The affective response model: A theoretical framework of affective concepts and their relationships in the ICT context. In MIS Quarterly, 37 (1) pp. 247-274*

*Zhang, Ping and Dran, Gisela M. von (2002): User Expectations and Rankings of Quality Factors, in Different Web Site Domains. In International Journal of Electronic Commerce, 6 (2) pp. 9-33*

*Zhang, Ping and Li, Na (2004): Love at first sight or sustained effect The role of perceived affective quality on users' cognitive reactions to IT. In: Proceedings of the 25th International Conference on Information Systems ICIS'04 December 12-15, 2004, Washington, USA.*

*Zheng, Jun, Veinott, Elizabeth S., Bos, Nathan, Olson, Judith S. and Olson, Gary M. (2002): Trust without touch: jumpstarting long-distance trust with initial social activities. In: Terveen, Loren (ed.) Proceedings of the ACM CHI 2002 Conference on Human Factors in Computing Systems Conference April 20-25, 2002, Minneapolis, Minnesota. pp. 141-146*

## About the author

### Dianne Cyr



Personal Homepage:
http://www.diannecyr.com

Dr. Dianne Cyr is a Professor in the Beedie School of Business at Simon Fraser University in Vancouver. She is the author of 4 books and over 100 research articles, book chapters or proceedings. Dr. Cyr has received numerous awards including Best paper for 2009 in the top ranked MISQ journal. My background is varied and interdisciplinary. Both my Bachelor and Masters degrees are in Psychology, and I worked in clinical psychology for the better part of a decade before returning to university to embark on doctoral studies. The earlier training held me in good stead for my Ph.D. research which was focused on the linkage of strategy and human resource management in international joint ventures. I am currently at Simon Fraser University as a Full Professor with tenure in Management Information Systems. Since 1994 my primary role has been as a university professor, coupled with activities in the business world. I joined Simon Fraser University (SFU) in 1994 as an Adjunct Professor, and carried a very full teaching load (of up to 8 courses in some years), while at the same time conducting research, publishing, and organizing a consulting practice. An opportunity arose for full-time academic employment at the start-up Technical University of British Columbia (TechBC) in 1998. As one of the founding faculty, I was involved in numerous activities to shape the mission, curriculum and structures for TechBC. More specifically, I developed the Management and Technology program until the closure of the university in 2002. At this time, TechBC was merged into SFU, and another academic chapter began. As the President of Global Alliance Management (1993-2004), I provided consulting and training services in the area of joint ventures and strategic alliances to small and mid-sized companies. During this time I made numerous public presentations on the topic of joint ventures or alliances, and developed a program in alliances for the British Columbia Trade Development Corporation. As a Director of Canada Sri Lanka Capital Corporation (1994-2003), I was involved in the development of an agro-industrial joint venture in Sri Lanka.

## Your notes and thoughts on chapter 40

Record your notes and thoughts on this chapter. If you want to share these thoughts with others online, go to the bottom of the page at:
http://www.interaction-design.org/encyclopedia/emotion_and_website_design.html

**NOTES:**

# Chapter 42

# Design 4 All

*by Constantine Stephanidis.*

## 42.1 WHAT IS DESIGN FOR ALL?

Contemporary interactive technologies and environments are used by a multitude of users with diverse characteristics, needs and requirements, including able-bodied and disabled people, people of all ages, people with different skills and levels of expertise, people from all over the world with different languages, cultures, education, etc. Additionally, interactive technologies are penetrating all aspects of everyday life, in communication, work and collaboration, health and well being, home control and automation, public services, learning and education, culture, travel, tourism and leisure, and many others. New technologies targeted to satisfy human needs in the above contexts proliferate, whether stationary or mobile, centralized or distributed, visible or encapsulated in the environment. A wide variety of devices is already available, and new ones tend to appear frequently and on a regular basis.

In this context above, interaction design acquires a new dimension, and the question arises of how it is possible to design systems that permit systematic and cost-effective approaches to accommodating all users. Design for All is an umbrella term for a wide range of design approaches, methods, techniques and tools to help address this huge diversity of needs and requirements in the design of interactive technologies. Design for All entails an effort to build access features into a product, starting from its conception and throughout the entire development life-cycle.

This chapter:

- introduces Design for All through a brief excursus into its roots and origin
- provides an overview of the dimensions of diversity which make Design for All a necessity in today's technological landscape
- presents the main perspectives on Design for All and the related technical approaches
- discusses some commonly design methods and techniques in the Design for All context
- presents both consolidated and emerging interaction technologies and techniques
- identifies future directions in the context of the emerging paradigm of Ambient Intelligence.

### 42.1.1 *Brief history*

The concept of Design for All was introduced in the Human-Computer Interaction (HCI) literature at the end of the nineties, following a series of research efforts mainly funded by the European Commission (Stephanidis and Emiliani, 1999;

Stephanidis et al., 1998; Stephanidis et al., 1999). Design for All in HCI is rooted in the fusion of three traditions:

i. user-centered design placing the user at the center of the interaction design process
ii. accessibility and assistive technologies for disabled people
iii. Universal Design for physical products and the built environment.

#### 42.1.1.1 From User-centered design to Design for All

User-centered design (Vrendenburg et al., 2001; ISO 1999; ISO 2010) is an approach to interactive system design and development that focuses specifically on making systems usable. It is an iterative process whose goal is the development of usable systems, achieved through the involvement of potential users during the design of the system.

**Definition**

User-centered design

An approach to designing ease of use into the total user experience with products and systems. It involves two fundamental elements—multidisciplinary teamwork and a set of specialized methods of acquiring user input and converting it into design.

Vrendenburg et al., 2001

User-centered design includes four iterative design activities, all involving direct user participation:

1. understand and specify the context of use, the nature of the users, their goals and tasks, and the environment in which the product will be used;

2. specify the user and organizational requirements in terms of effectiveness, efficiency and satisfaction; and the allocation of function between users and the system;
3. produce designs and prototypes of plausible solutions; and
4. carry out user-based assessment.

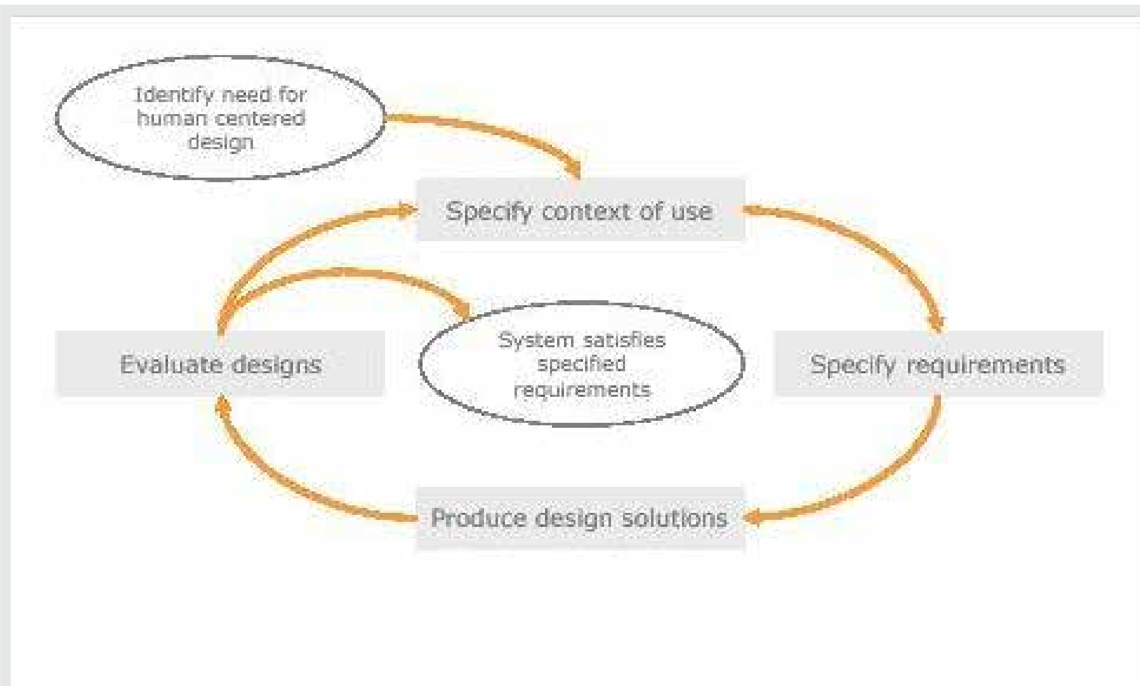





Figure 42.1: User-centered design cycle.

User-centered design requires:

- Active involvement of users and clear understanding of user and task requirements. The active involvement of end-users is one of the key strengths, as it conveys to designers the context of use in which the system will be used, potentially enhancing the acceptance of the final outcome.

- The appropriate allocation of functions between the user and the system. It is important to determine which aspects of a job or a task will be handled by users and which can be handled by the system itself. This division of labor should be based on an appreciation of human capabilities and their limitations, and on a thorough grasp of the particular demands of the task.
- Iteration of design solutions. Iterative design entails receiving feedback from end-users following their use of early design solutions. The users attempt to accomplish "real world" tasks using the prototype. The feedback from this exercise is used to further develop the design.
- Multi-disciplinary design teams. User centered system development is a collaborative process that benefits from the active involvement of various parties, each of whom have insights and expertise to share. Therefore, the development team should be made up of experts with technical skills in various phases of the design life cycle. The team might thus include managers, usability specialists, end-users, software engineers, graphic designers, interaction designers, training and support staff, and task experts.

User-centered design claims that the quality of use of a system, including usability and user health and safety, depends on the characteristics of the users, tasks, and the organizational and physical environment in which the system is used. Also, it stresses the importance of understanding and identifying the details of this context in order to guide early design decisions, and provides a basis for specifying the content in which usability should be evaluated. However, it is limited in its capability of addressing the diversity in user requirements, as it fosters the traditional perspective of "typical" or "average" users interacting with desktop computers in working environments (Stephanidis, 2001). While user-centered design focuses on maintaining a multidisciplinary and user-involving perspective into systems development, it does not specify how designers can cope with radically different user groups.

#### 42.1.1.2 From Accessibility and Assistive Technologies to Design for All

Accessibility in the context of HCI refers to the access by people with disabilities to Information and Communication Technologies (ICT). Interaction with ICT may be affected in various ways by the user's individual abilities or functional limitations that may be permanent, temporary, situational or contextual. For example, someone with limited visual functions will not be able to use an interactive system which only provides graphical output, while someone with limited bone or joint mobility or movement functions, affecting the upper limbs, will encounter difficulties in using an interactive system that only accepts input through the standard keyboard and mouse.

Accessibility in the context of HCI aims to overcome such barriers by making the interaction experience of people with diverse functional or contextual limitations as near as possible to that of people without such limitations.

The interaction process can be roughly analyzed as follows:

- the user provides input to the system through an action using an input device (e.g., the user pushes a mouse button to enter a command);
- the input is interpreted by the system (e.g., the system recognizes and executes the command);
- the system generates a response (e.g., the system generates a response message for the user notifying the execution of the command);

- the response is presented to the user through a system action using an output device (e.g., the message is displayed on the screen through a message window);
- the user perceives and interprets the response (e.g., the users sees the message in the message window on the screen and reads it).

A physical device is an artifact of the system that acquires (input device) or delivers (output device) information. Examples include keyboard, mouse, screen, and loudspeakers. An interaction technique involves the use of one or more physical devices to allow end-users provide input or receive output during the operation of a system.

An interaction style is a set of perceivable interaction elements used by the user (through an interaction technique) or the system to exchange information sharing common aesthetic and behavioral characteristics. In graphical user interfaces the term interaction style is used to denote a common look and feel among interaction elements. Typical examples are menu selection and direct manipulation. Interaction elements compose the user interface of a system with user interaction resulting from physical actions. A physical action is an action performed either by the system or the user on a physical device. Typically, system actions concern feedback and output, while user actions provide input. Examples of input actions include pushing a mouse button or typing on a keyboard. Different interaction techniques and styles exploit different sensory modalities. In practice, the modalities related to seeing and hearing are the most commonly employed for output, whereas haptics is less used. Interestingly, however, haptics remains the primary modality for input (e.g., typing, pointing, touching, sliding, grabbing, etc). Taste and smell have only recently started being investigated for output purposes.

In summary, the actual human functions involved in interaction are motion, perception and cognition. In this context, accessibility implies that:

- the input devices and related interaction techniques are such that their manipulation by users is feasible (i.e., they are compatible with the user's functions related to motion);
- the adopted interaction styles (and the resulting user interfaces) can be perceived by the users (i.e., they are compatible with the user's sensory functions);
- the adopted interaction styles (and the resulting user interfaces) can be understood by the users (i.e., they are compatible with the user's cognitive functioning).

Given the degree of human diversity as regards the involved functions, accessibility requires the availability of alternative devices and interaction styles to accommodate different needs.

In traditional efforts to improve accessibility, the main direction followed has been to enable disabled users to access interactive applications originally developed for able-bodied users through appropriate assistive technologies.

Assistive Technology (AT) is a generic term denoting a wide range of accessibility plug-ins including special-purpose input and output devices and the process used in selecting, locating, and using them. AT promotes greater independence for people with disabilities by enabling them to perform tasks that they were originally unable to accomplish, or had great difficulty accomplishing. In this context, it provides enhanced or alternative methods to interact with the technology involved in accomplishing such tasks.

## Definition

Assistive or Adaptive Technology commonly refers to "…products, devices or equipment, whether acquired commercially, modified or customized, that are used to maintain, increase or improve the functional capabilities of individuals with disabilities…"

Assistive Technology Act of 1998

Popular Assistive Technologies include screen readers and Braille displays for blind users, screen magnifiers for users with low vision, alternative input and output devices for motor impaired users (e.g., adapted keyboards, mouse emulators, joystick, binary switches), specialized browsers, and text prediction systems).



Figure 42.2: Adapted keyboard.



Figure 42.3: Footmouse.

Assistive Technologies are essentially reactive in nature. They provide product-level and platform-level adaptation of interactive applications and services originally designed and developed for able-bodied users. The need for more systematic and proactive approaches to the provision of accessibility has led to the concept of Design for All. Such shift from accessibility, as traditionally defined in the assistive technology sector, to a Design for All perspective, is due to: (i) the rapid pace of technological developments, with many new systems, devices, applications, and users, making accessibility add-ons very difficult to develop; (ii) an increased interest in people at risk of technological exclusion, not only limited to people with disabilities. Under this perspective, accessibility has to be designed as a primary system feature, rather than decided upon and implemented *a posteriori*, thus integrating accessibility into the design process of all applications and services.

#### 42.1.1.3 From Universal Design to Design for All in HCI

Proactive approaches toward addressing people's diversity first emerged in engineering disciplines, such as civil engineering and architecture, with many applications in interior design, building and road construction.

The term Universal Design was coined by the architect Ronald L. Mace to describe the concept of designing all products and the built environment to be both aesthetically pleasing and usable to the greatest extent possible by everyone, regardless of their age, ability, or status in life (Mace et al., 1991). Although the scope of the concept has always been broader, its focus has tended to be on the built environment.

**Definition**

**Universal Design:**
*is the design of products and environments to be usable by all people, to the greatest extent possible, without the need for adaptation or specialized design.*

*The intent of Universal Design is to simplify life for everyone by making products, communications, and the built environment more usable by as many people as possible at little or no extra cost. Universal Design benefits people of all ages and abilities.*

*(Mace et al., 1991)*

A classic example of Universal Design is the kerb cut (or sidewalk ramp), initially designed for wheelchair users to navigate from street to sidewalk, and today widely used in many buildings. Other examples are low-floor buses, cabinets with pull-out shelves, as well as kitchen counters at several heights to accommodate different tasks and postures.



Figure 42.4: Accessible home.



Figure 42.5: Accessible traffic light.

Perhaps the most common approach in Universal Design is to make information about an object or a building available through several modalities, such as Braille on elevator buttons, and acoustic feedback for traffic lights. People without disabilities can often benefit too. For example, subtitles on TV or multimedia content intended for the deaf can also be useful to non-native speakers of a language, to children for improving literacy skills, or to people watching TV in noisy environments.

The seven Principles of Universal Design were developed in 1997 by a working group of architects, product designers, engineers, and environmental design researchers, led by Ronald Mace at North Carolina State University. The Principles "may be applied to evaluate existing designs, guide the design process and educate both designers and consumers about the characteristics of more usable products and environments."

***Equitable use***: The design is useful and marketable to people with diverse abilities

***Flexibility in use***: The design accommodates a wide range of individual preferences and abilities

***Simple and intuitive***: Use of the design is easy to understand, regardless of the user's experience, knowledge, language skills, or current concentration level

***Perceptible information***: The design communicates necessary information effectively to the user, regardless of ambient conditions or the user's sensory abilities

***Tolerance for error***: The design minimizes hazards and the adverse consequences of accidental or unintended actions

***Low physical effort***: The design can be used efficiently and comfortably and with a minimum of fatigue

***Size and space for approach and use***: Appropriate size and space is provided for approach, reach, manipulation, and use regardless of user's body size, posture, or mobility

In the context of HCI, the above concepts were used by the end of the nineties to denote design for access by anyone, anywhere and at any time to interactive products and services.

The term *Design for All either subsumes, or is a synonym of, terms such as accessible design, inclusive design, barrier-free design, universal design, each highlighting different aspects of the concept.*

*(Stephanidis et al., 1998)*

**Definition**

***Design for All in HCI:***

- *Is the conscious and systematic effort to proactively apply principles and methods, and employ appropriate tools,*
- *in order to develop Information Technology &Telecommunications (IT&T) products and services which are accessible and usable by all citizens,*
- *thus avoiding the need for a posteriori adaptations, or specialized design.*

*This entails an effort to build access features into a product starting from its conception, throughout the entire development life-cycle.*

*(Stephanidis et al., 1998)*

Design for All in HCI implies a reconsideration of traditional design qualities such as accessibility and usability.

**Definitions**

***Accessibility:***

for each task a user has to accomplish through an interactive system, and taking into account specific functional limitations and abilities, as well as other relevant contextual factors, there is a sequence of input and output actions which leads to successful task accomplishment (Savidis and Stephanidis, 2004).

***Usability:***

capability of all supported paths towards task accomplishment to "maximally fit" individual users' needs and requirements in the particular context and situation of use (Savidis and Stephanidis, 2004).

*It follows that accessibility is a fundamental prerequisite of usability, since there may not be optimal interaction if there is no possibility of interaction in the first place.*

## 42.2 DIMENSIONS OF DIVERSITY

### 42.2.1 *User diversity*

The term "user diversity" refers to the various differences among users in their perception, manipulation, and utilization of technology. Understanding the various dimensions of user diversity helps design and develop user interfaces that maximize benefits for different types of users. There are several dimensions of diversity that differentiate people's interactions with technology.

#### 42.2.1.1 Disability

At the heart of accessibility lies a focus on human diversity in relation to access and use of ICTs. Main efforts in this direction are concerned with the identification and study of diverse target user groups (e.g., people with various types of disabilities and elderly people), as well as of their requirements for interaction, and of appropriate modalities, interactive devices and techniques to address their needs. Much experimental work has been conducted in recent years in order to collect and elaborate knowledge of how various disabilities affect interaction with technology. Such understanding can be facilitated by the functional approach of the "International Classification of Functioning, Disability and Health (ICF)" (World Health Organization 2001), where the term disability is used to denote a multidimensional phenomenon resulting from the interaction between people and their physical and social environment. This allows grouping and analysis of limitations that are not only due to impairments, but also, for example, due to environmental reasons.

#### 42.2.1.1.1 PERCEPTION

Visual and auditory impairments significantly affect human-computer interaction.

Blindness means anatomic and functional disturbances of the sense of vision of sufficient magnitude to cause total loss of light perception, while visual impairment refers to any deviation from the generally accepted norm. Blind users benefit from using the auditory and the haptic modality for output and input purposes. In practice, blind users' interaction is supported through screen readers (i.e., specialized software which reads aloud a graphical interface to the user (Asakawa and Leporini, 2009)), or speech output, and Braille displays (see section *Haptics*). The latter are haptic devices, but require knowledge of the Braille code. Audio (non-speech) sound can also be used to improve the blind user's interactive experience (Nees and Walker, 2009). Blind users can use keyboards and joysticks, though not the mouse. Therefore, all actions in a user interface must be available without the use of the mouse. It is also important for both output and input that the provided user interface is structured in such a way as to minimize the time required to access specific important elements (e.g., menus or links) when they are made available in a sequential fashion (e.g., through speech).

Less severe visual limitations are usually addressed by increasing the size of interactive artifacts, as well as color contrast between the background and foreground elements of a user interface. Specific combinations of colors may be necessary for users with various types of color-blindness.

Hearing impairments include any degree and type of auditory disorder, on a scale from slight to extreme. Hearing limitations can significantly affect interaction with technology. Familiar coping strategies for hearing-impaired people include the use of hearing aids, lip-reading and telecommunication devices for the deaf. Deaf users may not be comfortable with written language in user interfaces, and may benefit from sign-language translations of the information provided (see section *Sign Language*).

#### 42.2.1.1.2 MOTION

The nature and causes of physical impairments vary; however, the most common problems faced by individuals with physical impairments include poor muscle control, weakness and fatigue, difficulty in walking, talking, seeing, speaking, sensing or grasping (due to pain or weakness), difficulty reaching things, and difficulty doing complex or compound manipulations (push and turn). Individuals with severe physical impairments usually have to rely on assistive devices such as mobility aids, manipulation aids, communication aids, and computer interface aids (Keates, 2009).

Motion impairments interfere with the functions that are necessary for interacting with technology. For example, using a mouse and a keyboard can be challenging or painful. Therefore, motor-impaired users may benefit from specialized input devices minimizing movement and physical effort required for input, including adapted keyboards, mouse emulators, joystick and binary switches, often used in conjunction with an interaction technique called scanning (see section *Scanning-based Interaction*), as well as virtual on-screen keyboards for text input, sometimes supported by text prediction systems. Other interaction techniques which have been investigated for use by people with motion functions limitations are voice input of spoken commands (see section *Speech*), keyboard and mouse simulation through head movements (see section *Gestures and Head Tracking*), and eye-tracking (see section *Eye-tracking*). Brain-computer interfaces which allow control of an application simply by thought are also being investigated for supporting the communication of people with severe motor impairments (see section *Brain Interfaces*).

#### 42.2.1.1.3 COGNITION

Cognition is the ability of the human mind to process information, think, remember, reason, and make decisions. The extent of cognitive abilities varies greatly from person to person. Cognitive disability entails a substantial limitation in one's capacity to think, including: conceptualizing, planning, and sequencing thoughts and actions; remembering, interpreting subtle social cues, and understanding numbers and symbols. Cognitive disabilities can stem from brain injury, Alzheimers disease and dementia, severe and persistent mental illness, and stroke. Cognitive disabilities are many and diverse, individual differences are often very pronounced for these user groups, and it is particularly difficult to abstract and generalize the issues involved in researching user requirements for this part of the population.

Various cognitive limitations and learning difficulties may affect the interaction process. General principles which facilitate access for users with some types of cognitive difficulties, but also for other user groups, such as, for example, older users, are to keep the user interface as simple and minimalistic as possible, provide syntactically and lexically simple text, reduce the need to rely on memory, allow sufficient time for interaction, and support user attention (Lewis, 2009). Specific developmental learning conditions such as dyslexia also require particular care in the use of text, fonts, colors, contrast and images in order to facilitate comprehension.

#### 42.2.1.2 AGE

Age plays a significant role in how a person perceives and processes information. Knowing the age of the target population of a technology product can provide vital clues about how to present information, feedback, video, audio, etc. Two user groups have particular requirements dependent on age: children (defined as users below the age of 18, with particular focus on children under the age of 12) and older persons (usually defined as users over the age of 65).

##### 42.2.1.2.1 CHILDREN

In the United States, nearly half (48%) of all children aged six and under have used a computer, and more than one in four (30%) have played video games. By the time they are in the four-to six-year-old range, seven out of ten have used a computer (Wartella et al., 2005).

The emergence of children as an important new user group of technology dictates the importance of supporting them in a way that is useful, effective, and meaningful for their needs.

The physical and cognitive abilities of children develop over a period of years from infancy to adulthood. Children, particularly those who are very young, do not have a wide repertoire of experiences that guide their responses to cues. In addition to this lack of experience, children perceive the world differently from adults, and have their own likes, dislikes, curiosities and needs that are different from adults. Therefore, children should be regarded as a different user population with its own culture and norms (Bruckman and Bandlow, 2002).

The design of applications for children poses a special challenge, as designers must learn how to perceive systems through the eyes of a child. For example, audio feedback may alarm very young children and extremely bright colors and video could easily distract them from the task.

#### 42.2.1.2.2 OLDER USERS

There is overwhelming evidence that the population of the developed world is ageing. In addition to the growing population of elders in the United States (20% of the entire population by 2030), these numbers are increasing on the global scale as well. It is estimated that, for the first time in history, the population of older adults will exceed the population of children (age 0-14) in the year 2050. Almost 2 billion people will be considered older adults by 2050 (US Department of Health and Human Services Administration on Ageing 2009).

This large and diverse user group, with a variety of different physical, sensory, and cognitive capabilities, can benefit from technological applications which can enable them to retain their health, well being and independent living.

Older users may experience a decrease in motor, sensory and cognitive functioning, which may lead to combined impairments and highly affect interaction (Kurniawan, 2009). Principles for providing accessibility to older users include improved contrast, enlargement of information presented on the screen, careful organization of information, choice of appropriate input devices, avoiding relying on memory, and design simplicity.

Older people have a vast set of memories from experiences in the past that compose a large repertoire. This naturally influences their feelings towards technology. Older users may feel a sense of resistance to certain technologies, especially when dealing with applications for tasks that people are used to completing without technology, such as online banking systems. The feeling of being “forced” to adapt to technology during the later years of life can add to these feelings of resistance.

### 42.2.1.3 Computer use expertise

The wide use of technology by a large group of the population has resulted in increased comfort with basic technological tools. However, the level of comfort and the ease of use of technology vary significantly depending on the skill levels of users (Ashok and Jacko, 2009).

Some groups of users are unfamiliar with technology, particularly older users and those with minimal or no education, but are nevertheless required to use computing tools in order to keep up with the current evolution of society. The result is a mix of users with great diversity in technology skill level.

The challenge of designing systems for users who fall within a wide and uneven spectrum of skills can be daunting. This is especially so because designers are typically experts in their respective domains and find it difficult to understand and incorporate the needs of novices. Judging the skill levels of users can be more difficult than assessing impairments or difficulties because users who are experts on a particular tool may find a new replacement tool hard to use and understand. This results in a situation where a person who apparently is an expert actually behaves like a novice. Feedback from users with different skill levels can provide fresh perspectives and new insights.

Including useful help options, and explanations which can be expanded and viewed in more detail, consistent naming conventions, and uncluttered user interfaces, are just a few ways in which technology can be made accessible by users with less knowledge of the domain and system without at the same time reducing efficiency for expert users. In fact, these suggestions are guidelines of good design, which will benefit all users, irrespective of skill level.

#### 42.2.1.4 Culture and Language

In today's world, due to globalized technology, there is a significant shift in the perception, understanding, and experience of culture. The inclusion of this knowledge in technology will lead to more inclusiveness and tolerance.

Language is an integral part of culture and much can be lost in translation due to language barriers. For example, many technological applications use English and this in itself could be a restricting factor for people who do not speak or write the language. Abbreviations, spelling, and punctuation are all linguistic variables. The connection between language and the layout of text on technical applications is a factor to be considered, since certain languages such as English and French lend themselves to shorter representations, while other languages may require longer formats.

The term “localization” refers to customizing products for specific markets to enable effective use. Included in localization are language translations, changes to graphics, icons, content, etc.

Other differences in culture include interpretations of symbols, colors, gestures, etc. For example, there are differences in the use of colors (green is a sacred color in Islam, yellow in Buddhism) and in the reading direction (e.g. left to right in N. America and Europe, right to left in the Middle East). Ideas on clothing, food, and aesthetic appeal also vary from culture to culture.

These numerous differences make it imperative that designers are sensitive to these differences during the creation of technology and avoid treating all cultures as the same. Rather than neutralize cultural and linguistic differences, Design for All acknowledges, recognizes, appreciates, and integrates these differences (Marcus and Rau, 2009).

#### 42.2.1.5 Social Issues

Globalization has created an environment of rich information and easy communication. Social issues such as economic and social status pose a serious challenge with respect to access to technology. In many parts of the world, only the wealthier segments of society have the opportunity to use technology and benefit from it.

Poverty, social status and limited educational opportunities create barriers to technology access. Designing applications that are equally accessible and equally easy to use for every single socio-economic group in the world is virtually impossible, but there are lessons to be learned from considering the needs of various social groups.

Studies have revealed that a certain level of education, technical education to be precise, is required to receive optimal productivity from the use of technology (Castells, 1999). The realization that technological benefits are available more readily to the educated conveys a simple message regarding the responsibility of designers, developers, engineers, and all those involved in the creation of technology. This team of people creates and distributes technology, and it is critically important for them to be educated in matters of universal access and issues related to the diversity of users, including the need to consider designing for the under-educated. Designing for technological literacy becomes therefore a top priority.

### 42.2.2 *Diversity in the technological environment*

Diversity does not only concern users, but also interaction environments and technologies, which are continuously developing and diversifying.

Temporary states of impairment may be created by the particular contexts in which users interact with technology. For instance, a working environment in which noise level and visual distractions of the environment are high can interfere with the efficient use of computer-based applications. Impairments caused by contextual factors are known as situationally-induced impairments (Sears et al., 2003). Technology itself can also cause situationally-induced impairments. For example, when screens are too small, the user may become vision-impaired in this particular situation.

The diffusion of the Internet and the proliferation of advanced interaction technologies (e.g., mobile devices, network attachable equipment, virtual reality, agents, etc.) signify that many applications and services are no longer limited to the visual desktop, but span over new realities and interaction environments.

Overall, a wide variety of technological paradigms plays a significant role for Design for All, either by providing new interaction platforms, or by contributing at various levels to ensure and widen access.

The World Wide Web and its technologies are certainly a fundamental component of this landscape. Various challenges exist and solutions have been elaborated to make the Web accessible to all. The World Wide Web offers much for those who are able to access its content, but at the same time access is limited by serious barriers due to limitations of visual, motor, language, or cognitive abilities.

Another very important and rapidly progressing technological advance is that of mobile computing. Mobile devices acquire an increasingly important role in everyday life, both as dedicated tools, such as media players, and multi-purpose devices, such as smart phones. The device needs to be easy to use, even on the move. Mobile interaction often brings forward contradictory design goals and requirements. The environments of mobile contexts are demanding due to characteristics such as noise or poor lighting. The user may need to multitask and that leaves only part of his/her attention for using the device. Also cultural differences and user expectations have a major impact on the use of the devices. These characteristics of mobile devices and usage situations set high demands for design. Similar to any other application field, mobile user groups can be defined by age, abilities and familiarity with the environment. However, the requirements that each user group sets for mobile devices and services vary in different contexts of use.

Mobility brings about the challenge that contexts of use vary greatly and may change, even in the middle of usage situations. The variable usage contexts need to be taken into account when designing mobile devices and services. The initial assumption of mobile devices and services is that they can be used "anywhere". This assumption may not always be correct; the environment and the context create challenges in use. Using mobile phones is prohibited in some environments and in some places there may not be network coverage.

The user can be physically or temporally disabled. In dark or bright environments it may be hard to see the user interface elements. In a crowded place it may be difficult to carry on a voice conversation over the phone—even more difficult than in a face to face situation, when you can use non-verbal cues to figure out what the other person is saying if you cannot hear every word. In social communication the context plays an important role; people often start telephone conversations by asking the other person's physical location— "Where are you"—to figure out whether the context of the other person allows the phone call.

## 42.3 PERSPECTIVES AND APPROACHES

In the context of Design for All, user interface design methodologies, techniques and tools acquire increased importance. Various methods, techniques, and codes of practice have been proposed to enable authors proactively to take into account and appropriately address diversity in the design of interactive artifacts. Three fundamental approaches are outlined below.

### 42.3.1 *Guidelines and Standards*

Guidelines and standards have been formulated in the context of international collaborative initiatives towards supporting the design of interactive products, services and applications, which are accessible to most potential users without any modifications. This approach has been mainly applied to the accessibility of the World Wide Web.

#### 42.3.1.1 Web accessibility

> *Web accessibility means that people with disabilities can use the Web. More specifically, Web accessibility means that people with disabilities can perceive, understand, navigate, and interact with the Web, and that they can contribute to the Web. Web accessibility also benefits others, including older people with changing abilities due to aging.*
>
> http://www.w3.org/WAI/intro/accessibility.php

A number of accessibility guidelines collections have been developed (Pernice and Nielsen, 2001; Vanderheiden et al., 1996). In particular, the Web Content Accessibility Guidelines (WCAG) explain how to make Web content accessible to people with disabilities (W3C, 1999). Web "content" generally refers to the information in a Web page or Web application, including text, images, forms, sounds, etc. WCAG 1.0 provides 14 guidelines that are general principles of accessible design. Each guideline has one or more checkpoints that explain how the guideline applies in a specific area. WCAG foresees 3 levels of compliance, A, AA and AAA. Each level requires a stricter set of conformance guidelines, such as different versions of HTML (Transitional vs. Strict) and other techniques that need to be incorporated into code before accomplishing validation. Further to WCAG 1.0, in December 2008, the W3C published a new version of the guidelines, targeted to help Web designers and developers to create sites that better meet the needs of older users and users with disabilities. Drawing on extensive experience and community feedback, WCAG 2.0 (W3C, 2008) improves upon WCAG 1.0 and applies to more advanced technologies. Guidelines are also available for the usability of web interfaces on mobile devices (W3C, 2005). WCAG 2.0 guidelines are organized around four principles that provide the foundation for Web accessibility, namely *perceivable, operable, understandable, and robust*. The 12 guidelines provide the basic goals that authors should work toward in order to make content more accessible to users with different disabilities. The guidelines are not testable, but provide the framework and overall objectives to help authors understand the success criteria and better implement the techniques. For each guideline, testable success criteria are provided to allow WCAG 2.0 to be used where requirements and conformance testing are necessary such as in design specification, purchasing, regulation, and contractual agreements. In order to meet the needs of different groups and different situations, three levels of conformance are defined: A (lowest), AA, and AAA (highest). WACAG guidelines also address content used on mobile devices.

The use of guidelines is today the most widely adopted process by web authors for creating accessible web content. This approach has proven valuable for bridging a number of barriers faced today by people with disabilities. In addition, guidelines serve those with low levels of experience with computers, and facilitate interoperability with new and emerging technology solutions (e.g., navigator with voice recognition for car drivers). Additionally, guidelines constitute *de facto* standards, as well as the basis for legislation and regulation related to accessibility in many countries (Kemppainen, 2009). For example, the US government Section 508 of the US Rehabilitation Act (Rehabilitation Act of 1973, Amendment of 1998) provides a comprehensive set of rules designed to help web designers make their sites accessible.

However, many limitations arise in the use of guidelines for a number of reasons. These include the difficulty in interpreting and applying guidelines, which require extensive training. Additionally, the process of using, or testing conformance to, widely accepted accessibility guidelines is complex and time consuming. To address this issue, several tools have been developed enabling the semi-automatic checking of html documents. Such tools make the development of accessible web content easier, particularly since the checking of conformance does not rely solely on the expertise of developers. Developers with limited experience in web accessibility can use such tools for evaluating web content and without the need to go through a large number of checklists.

Despite the proven usefulness of WCAG for web accessibility, it is common for web content manufacturers to ignore or overlook them, thus limiting the ability of disabled users to navigate through the information and services offered by a website. The guideline principles are therefore far from being well integrated, even to public Web sites where legislation enforces it. Recent studies reveal that web accessibility metrics are worsening worldwide (Basdekis, et al., 2010).

As a final consideration, guidelines are usually applied following a "one-size-fits-all" approach to accessibility, which, while ensuring a basic level of accessibility for users with various types of disabilities, does not support personalization and improved interaction experience.

#### 42.3.1.2 Other Accessibility Guidelines and Standards

Besides web accessibility, guidelines and standards are available also for other types of applications. For example, major software companies have launched accessibility initiatives and provide accessibility guidelines for developers using their tools and development environments. Examples are (Microsoft, 2013), (Adobe, 2013), and (IBM, 2013).

Accessibility of multimedia content is also addressed by international consortia, especially in the domain of education, e.g., (IMS, 2013), but also by content providers, e.g., (BBC, 2013).

Other accessibility related activities by international, European and National standardization bodies are discussed in details in (Engelen, 2009).

### 42.3.2 *User Interface Adaptation*

In the light of the above, it appears that design approaches focusing on the delivery of single interaction elements to be used by everybody offer limited possibilities of addressing the diverse requirements reflected in all users. Therefore, a critical property of interactive elements becomes their capability for some form of automatic adaptation and personalization (Stephanidis, 2001).

Adaption-based approaches promote the design of products which are easily adaptable to different users by incorporating adaptable or customizable user interfaces. This entails an effort to build access features into a product, starting from its conception and continuing throughout the entire development life-cycle.

Methods and techniques for user interface adaptation meet significant success in modern interfaces. Some popular examples include the desktop adaptations in various versions of Microsoft Windows, offering, for example, the ability to hide or delete unused desktop items, and personalization features of the desktop based on the preferences of the user, by adding helpful animations, transparent glass menu bars, live thumbnail previews of open programs and desktop gadgets (like clocks, calendars, weather forecast, etc.). Similarly, Microsoft Office applications offer several customizations, such as toolbars positioning and showing/hiding recently used options. However, adaptations integrated into commercial systems need to be set manually, and mainly focus on aesthetic preferences. In terms of accessibility and usability, only a limited number of adaptations are available, such as keyboard shortcuts, size and zoom options, changing color and sound settings, automated tasks, etc.

Research efforts in the past decades have elaborated more comprehensive and systematic approaches to user interface adaptations in the context of Design for All. The Unified User Interfaces methodology was conceived and applied (Savidis and Stephanidis, 2009) as a vehicle to ensure, efficiently and effectively through automatic adaptation, the accessibility and usability of UIs to users with diverse characteristics, also supporting technological platform independence, metaphor independence, and user-profile independence. Automatic UI adaptation seeks to minimize the need for *a posteriori* adaptations and deliver products that can be adapted for use by the widest possible end user population (adaptable user interfaces). This implies the provision of alternative interface instances depending on the abilities, requirements, and preferences of the target user groups, as well as the characteristics of the context of use (e.g., technological platform, physical environment). The main objective is to ensure that each end-user is provided with the most appropriate interactive experience at run-time.

Designing for automatic adaptation is a complex process. Designers should be prepared to cope with large design spaces to accommodate design constraints posed by diversity in the target user population and the emerging contexts of use. Therefore, designers need accessibility knowledge and expertise. Moreover, user adaptation must be carefully planned, designed and accommodated into the life-cycle of an interactive system, from the early exploratory phases of design, through to evaluation, implementation, deployment, and maintenance.

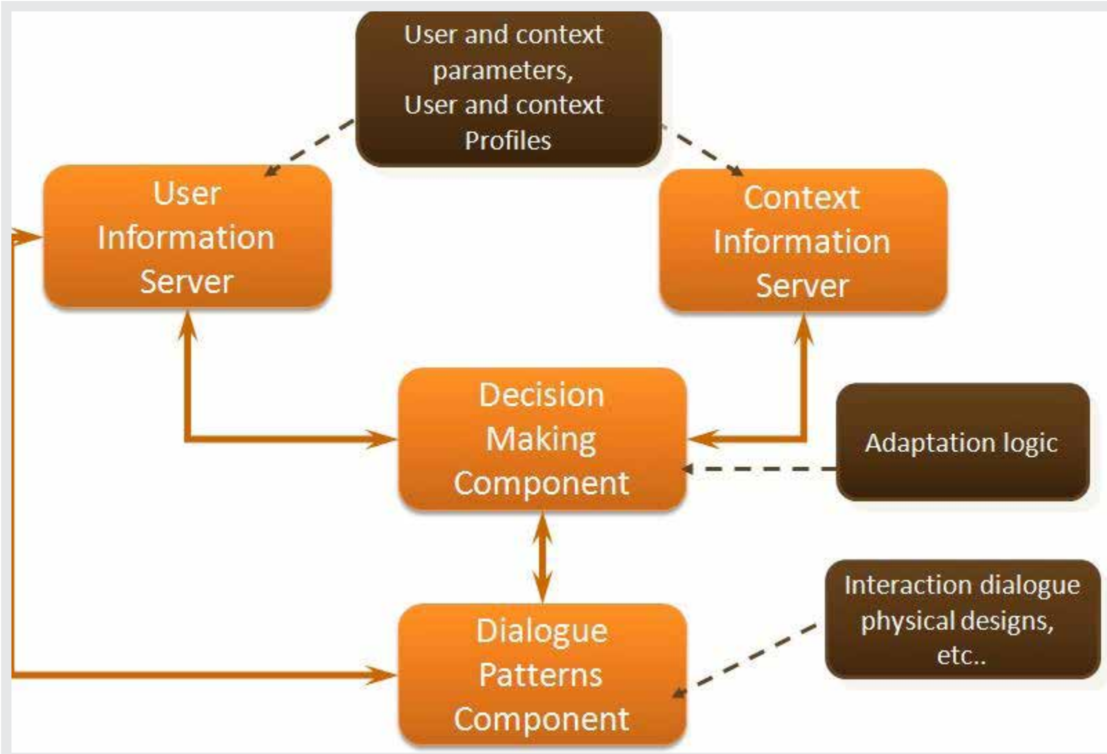





FIGURE 42.6: The architecture of an adaptable user interface.

A series of tools and components have been developed to support Unified User Interface design, including toolkits of adaptable interaction objects, languages for user profiling and adaptation decision making, and prototyping tools (Stephanidis et al., 2012). These tools are targeted to support the design and development of user interfaces capable of adaptation behavior, and more particularly the conduct and application of the Unified User Interface development approach. Over the years, these tools have demonstrated the technical feasibility of the approach and have contributed to reducing the practice gap between traditional user interface design and design for adaptation. They have been applied in a number of pilot applications and case studies.

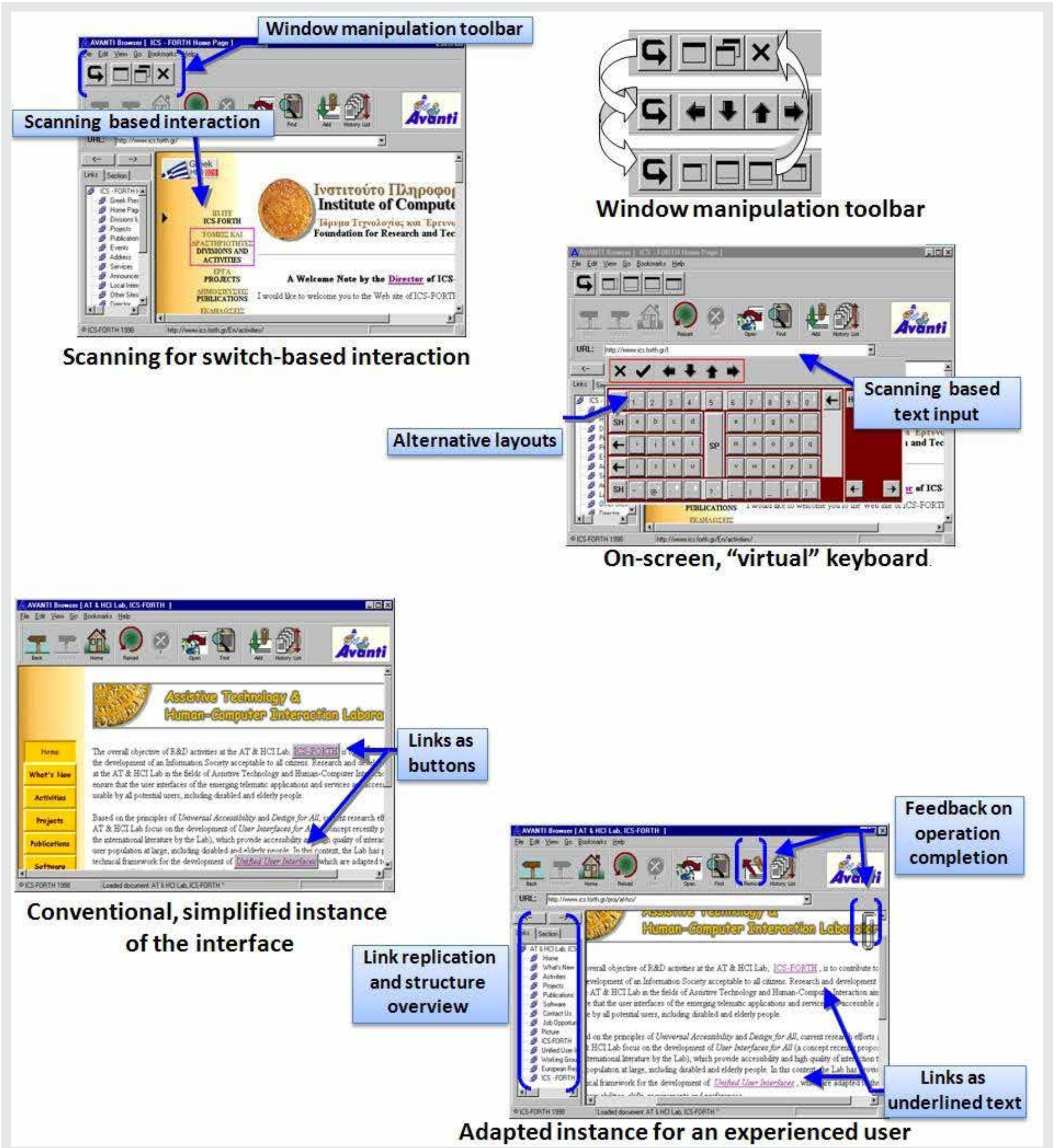





FIGURE 42.7: ***AVANTI browser***: a universally accessible web browser with a unified user interface. The AVANTI browser provides an interface to web-based information systems for a range of user categories, including: (i) "able-bodied" people; (ii) blind people; and (iii) motor-impaired people with different degrees of difficulty in employing traditional input devices. Stephanidis et al., 2001.

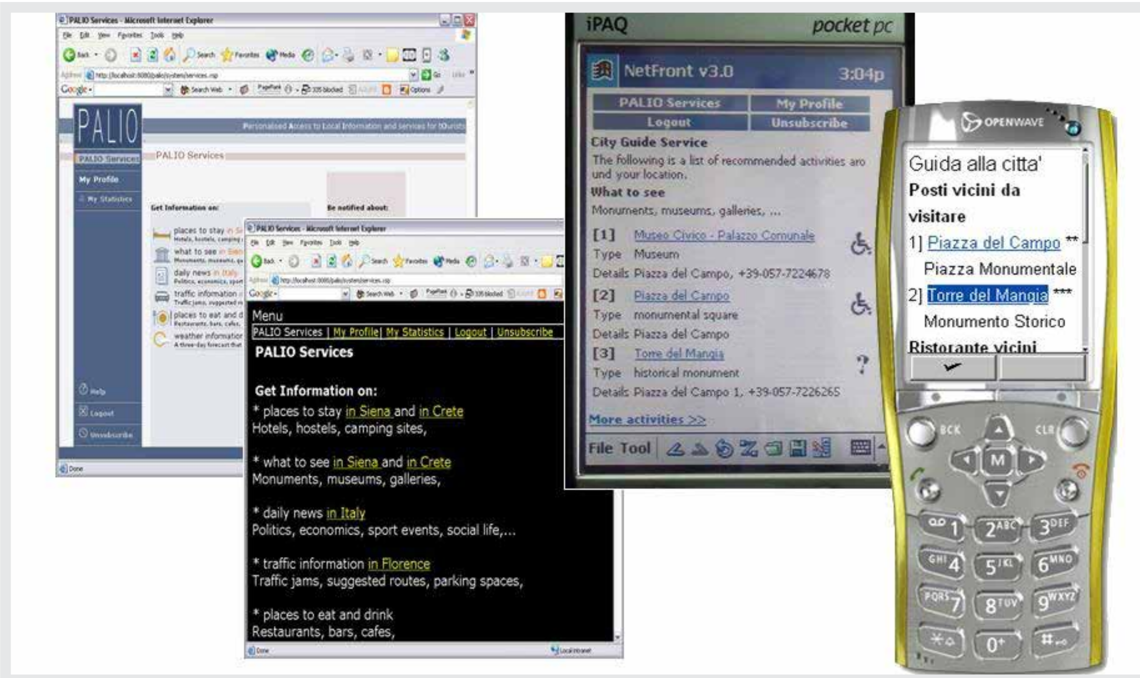





Figure 42.8: **PALIO**: a system that supports the provision of web-based services exhibiting automatic adaptation behavior based on user and context characteristics, as well as the user current location. Stephanidis et al., 2005.




FIGURE 42.9: ***UA-Chess***: a universally accessible multi-modal chess game, which can be played between two players, including people with disabilities (low-vision, blind and hand-motor impaired), either locally on the same computer, or remotely over the Internet. Grammenos et al., 2005.

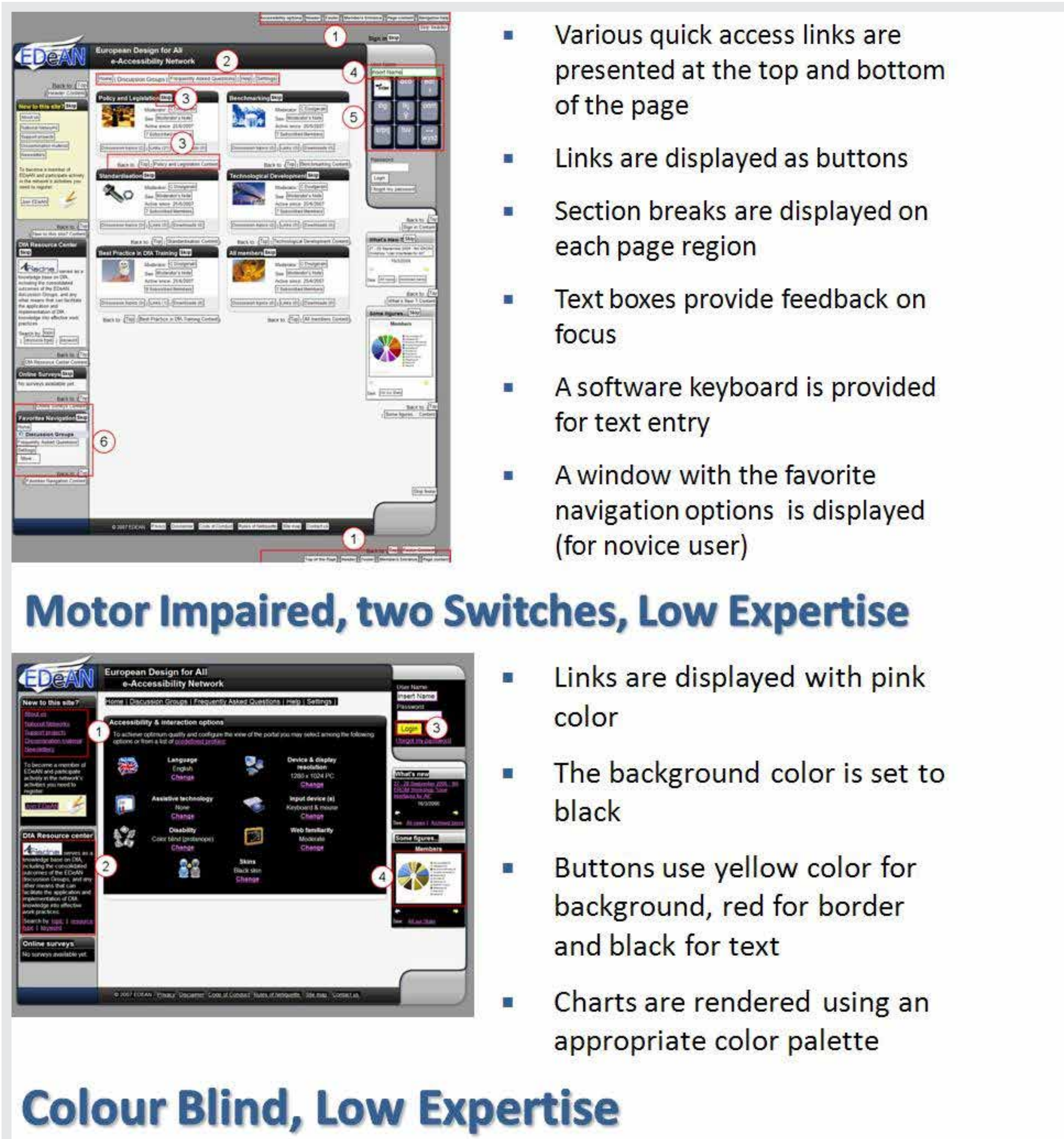





Figure 42.10: ***EDeAN Portal***: an adaptable web portal for the support of the activities of the EDeAN Network. Adaptations are performed server-side using a toolkit of adaptable interaction objects. Doulgeraki et al., 2009.

### 42.3.3 *Accessibility in the cloud*

The emergence of cloud computing in recent years is opening new opportunities for the provision of accessibility. Accessibility in the cloud is another approach targeted to move away from the concept of special "assistive technologies" and "disability access features" and towards providing more mainstream interface options for everyone, i.e., interfaces appropriate for people facing barriers in the use of interactive technologies due to disability, literacy or age-related issues, but also for people who just want a simpler interface, have a temporary disability, want access when their eyes are busy doing something else, want to rest their hands or eyes, or want to access information in noisy environments (Vanderheiden et al., 2013). Thus, the approach is targeted to everybody, including people with specific disabilities, such as (indicatively) blind and low-vision users, motor impaired users, users with cognitive impairments, hearing impairment users, and speech impaired users.

The main objective is to create an infrastructure that will support the creation of solutions that correspond to and respect the full range of human diversity. New systems need to allow prospective users to access and use solutions not just on a single computer, but on all of the different computers and ICT that they must use (at home, at work, when travelling, etc).

The infrastructure will enable users to declare requirements in functional terms (whether or not they fill into traditional disability categories) and allow service providers, crowd sourcing mechanisms, and commercial entities to respond to these requirements. This will mean that users with disabilities do not need to be constrained by their diagnostic categories, thus avoiding stereotyping and recognizing that everyone's requirements are different and that each individual's requirements may change according to the context.

Technically the approach is based on the creation of an explicit and implicit user profile (stored either locally or in the cloud), that automatically matches mainstream products and services with necessary access features and configures them according to the user's preferences and context of use. The infrastructure would consist of enhancements to platform and network technologies to simplify the development, delivery and support of access technologies and provide users with a way to instantly apply the access techniques and technologies they need, automatically, on any computers or other ICT.

Currently, the infrastructure is in its conceptualization and development phases. Although its basic concepts are based on past work and implementations, current efforts attempt to apply them together across the different platforms and technologies, and in a manner that can scale to meet the need. A wide range of delivery options is currently under development and testing, including auto-personalization of different OSs, browsers, phones, web apps, kiosks, ITMs, DTVs, smart homes and assistive technologies (cloud and installed).

## 42.4 DESIGN METHODS AND TECHNIQUES

The emergence of user-centered design (see section ***Brief History***) has led to the development and practice of a wide variety of design methods and techniques, mostly originating from the social sciences, psychology, organizational theory, creativity and arts, as well as from practical experience (Maguire and Bevan 2002). Many of these techniques are based on the direct participation of users or user representatives in the design process. However, the vast majority of available techniques have been developed with the "average" able-bodied user and the working environment in mind.

In Design for All, this precondition no longer holds, and the basic design principle of "knowing the user" becomes "knowing the diversity of users". Therefore, the issue arises of which methods and techniques can be fruitfully employed while addressing diversity in design, and how such techniques need to be used, revised and modified to optimally achieve this purpose. This is further complicated by the fact that, in Design for All, more than one group of users with diverse characteristics and requirements need to be taken into account (Antona et al., 2009).

Practical and organizational aspects of the involvement process play an important role when non-traditional user groups are addressed, and are critical to the success of the entire effort. Their importance should not be underestimated.

Very few design methods can be used as they stand when addressing diverse user groups. One of the main issues is therefore how to appropriately adapt and fine-tune methods to the characteristics of the people involved. Methods are mostly based on communication between users and other stakeholders in the design process. Therefore, the communication abilities of the involved users should be a primary concern.

### 42.4.1 *Observation*

Popular methods of exploring the user experience come from field-research in anthropology, ethnography and ethnomethodology (Beyer and Holtzblatt, 1997).

Direct observation is one of the hallmark methods of ethnographic approaches. It involves an investigator viewing users as they conduct some activity. The goal of field observation is to gain insight into the user experience as experienced and understood within the context(s) of use.

Examining the users in context is claimed to produce a richer understanding of the relationships between preference, behavior, problems and values.

Observation sessions are usually video-recorded, and the videos are subsequently analyzed. The effectiveness of observation and other ethnographic techniques can vary, as users have a tendency to adjust the way they perform tasks when knowingly being watched. The observer needs to be unobtrusive during the session and only pose questions when clarification is necessary.

Field studies and direct observation can be used when designing for users with disabilities and older users. This method does not specifically rely on the participants' communication abilities, and is therefore useful when the users have cognitive disabilities. Observational studies have been conducted with blind users in order to develop design insights for enhancing interactions between a blind person and everyday technological artifacts found in their home such as wristwatches, cell phones or software applications (Shinohara, 2006). Analyzing situations where work-arounds compensate for task failures reveals important insights for future artifact design for the blind, such as tactile and audio feedback, and facilitation of user independence.

A difficulty with direct observation studies is that they may in some cases be perceived as a form of "invasion" of the user's space and privacy, and therefore may not be well accepted, for example, by disabled or older people who are not keen to reveal their problems in everyday activities.

### 42.4.2 *Surveys and Questionnaires*

User surveys, originating from social science research, involve administering a set of written questions to a sample population of users, and are usually targeted to obtaining statistically relevant results. Questionnaires are widely used in HCI, especially in the early design phases but also for evaluation. Questionnaires need to be carefully designed in order to obtain meaningful results (Oppenheim, 2000).

Research shows that there are age differences in the way older and younger people respond to questionnaires. For example, older people tend to use the "Don't know" response more often than younger people. They also seem to use this answer when faced with questions that are complex in syntax. Their responses also seem to avoid the extreme ends of ranges. Having the researcher administer the questionnaire directly to the user may help to retrieve more useful and insightful information (Eisma et al., 2004). In-home interviews are effective in producing a wealth of information from the user that could not have been obtained by answering a questionnaire alone (Dickinson et al., 2002).

Since questionnaires and surveys address a wide public, and it is not always possible to be aware of the exact user characteristics (i.e., if they use Braille or if they are familiar with computers and assistive hardware and software), they should be available either in alternative formats or in accessible electronic form. The simple and comprehensible formulation of questions is vital. Questions must also be focused to avoid gathering large amounts of irrelevant information.

### 42.4.3 *Interviews*

Interviews are another ethnographically-inspired user requirements collection method. In HCI, as generally in software system development, it is a commonly used technique where users, stakeholders and domain experts are questioned to obtain information about their needs or requirements in relation to a system (Macaulay, 1996). Interviews can be unstructured (i.e., no specific sequence of questions is followed), structured (i.e., questions are prepared and ordered in advance) or semi-structured (i.e., based on a series of fixed questions with scope for the user to expand on their responses). The selection of representative users to be interviewed is important in order to obtain useful results. Interviews on a customer site by representatives from the system development team can be very informative.

Semi-structured interviews have been used to identify accessibility issues in mobile phones for blind and motor impaired users (Smith-Jackson et al., 2003. With older people, interviews as a means for gathering user requirements has also proven to be an effective method, although in-house interviews can be even more productive, because they tend to lead to spontaneous excursions into users' own experiences, and demonstrations of various personal devices used.

Obviously, interviews present difficulties when deaf people are involved, and sign-language translation may be necessary. Interviews are often avoided when the target user group is composed of cognitively and communication impaired people. Recently, a trend to conduct interviews on-line using chat tools has also emerged. An obvious consideration in this respect is that the used chat tool must be accessible and compatible with screen readers.

### 42.4.4 *Activity Diaries and Cultural Probes*

Diary keeping is another ethnographically inspired method which provides a self-reported record of user behavior over a period of time (Gaver et al., 1999). Participants are required to record activities they are engaged in during a normal day. Diaries allow identifying patterns of behavior that would not be recognizable through short-term observation. However, they require careful design and prompting if they are to be employed properly by participants. Diaries can be textual, but also visual, employing pictures and videos. Generalizing the concept of diaries, "cultural probes" are based on "kits" containing a camera, voice recorder, a diary, postcards and other items. They have been successfully employed for user requirements elicitation in home settings with sensitive user groups, such as former psychiatric patients and the elderly (Crabtree et al., 2003). Reading and writing a paper-based diary may be a difficult process for the blind and users with motor impairments. Therefore, diaries in electronic forms or audio recorded diaries should be used in these cases.

### 42.4.5 *Group discussions*

Brainstorming, originating from early approaches to group creativity, is a process where participants from different stakeholder groups engage in informal discussion to rapidly generate as many ideas as possible. All ideas are recorded, and criticism of others' ideas is forbidden. Overall, brainstorming can be considered as appropriate when the users to be involved have good communication abilities and skills (not necessarily verbal), but can also be adapted to the needs of other groups. This may have implications in terms of the pace of the discussion and generation of ideas.

Focus groups are inspired from market research techniques. They bring together a cross-section of stakeholders in a discussion group format. The general idea is that each participant can act to stimulate ideas, and that by a process of discussion, a collective view is established (Bruseberg and McDonagh-Philp, 2001). Focus groups typically involve six to twelve persons, guided by a facilitator. Several discussion sessions may be organized.

The main advantage of using focus groups for users with disabilities is that it does not discriminate against people who cannot read or write and it can encourage participation from people reluctant to be interviewed on their own or who feel they have nothing to say. During focus groups, various materials can be used for review, such as storyboards (see section *Scenario, Storyboards and Personas*).

This method should not be employed for requirements elicitation if the target user group has severe communication problems. Moreover, it is important that the discussion leader manages the discussion effectively and efficiently, allowing all users to participate actively in the process, regardless of their disability.

Focus groups have been used for eliciting expectations and needs from the learning disabled, as it was felt they would result in the maximum amount of quality data (Hall and Mallalieu, 2003). They allow a range of perspectives to be gathered in a short time period in an encouraging and enjoyable way. This is important, as, typically, people with learning disabilities have a low attention span.

Concerning older people, related research has found that it is not easy to keep a focus group of older people focused on the subject being discussed (Newell et al., 2007). Participants tend to drift their discussions off the subject matter as for them the focus group meeting is a chance to socialize. Thus, it is important to provide a social gathering as part of the experience of working with IT researchers rather than treat them simply as participants.

### 42.4.6 *Empathic modeling*

Empathic modeling is a technique intended to help designers/developers put themselves in the position of a disabled user, usually through disability simulation (Nicolle and Maguire, 2003). This technique was first applied to simulate age-related visual changes in a variety of everyday environmental tasks, with a view to eliciting the design requirements of the visually impaired in different architectural environments. Empathic modeling can be characterized as an informal technique, and there are no specific guidelines on how to use it.

- Modeling techniques for specific disabilities can be applied through simple equipment.
- Visual impairment due to cataracts can be simulated with the use of an old pair of glasses smeared with Vaseline.
- Total blindness is easy to simulate using a scarf or a bandage tied over the eyes.
- Total hearing loss can be easily simulated using earplugs.
- Upper limb mobility impairments can be simulated with the use of elastic bands and splints.

### 42.4.7 *Scenario, Storyboards and Personas*

Scenarios are widely used in requirements elicitation and, as the name suggests, are narrative descriptions of interactive processes, including user and system actions and dialogue. Scenarios give detailed realistic examples of how users may carry out their tasks in a specified context with the future system. The primary aim of scenario building is to provide examples of future use as an aid to understanding and clarifying user requirements and to provide a basis for later usability testing. Scenarios can help identify usability targets and likely task completion times (Carroll, 1995).

Storyboards are graphical depiction of scenarios, presenting sequences of images that show the relationship between user actions or inputs and system outputs. Storyboarding originated in the film, television and animation industry. A typical storyboard contains a number of images depicting features such as menus, dialogue boxes, and windows (Truong et al., 2006).

Another scenario-related method is called personas (Cooper, 1999), where a model of the user is created with a name, personality and picture, to represent each of the most important user groups.

The persona model is an archetypical representation of real or potential users. It is not a description of a real user or an average user. The persona represents patterns of users' goals and behavior, compiled in a fictional description of a single individual. Potential design solutions can then be evaluated against the needs of particular personas and the tasks they are expected to perform.

Zimmermann and Vanderheiden (2008) propose a methodology based on the use of scenarios and personas to capture the accessibility requirements of older people and people with disabilities and structure accessibility design guidelines. The underlying rationale is that the use of these methods has great potential to make this type of requirement more concrete and comprehensible for designers and developers who are not familiar with accessibility issues.

However, really reliable and representative personas can take a long time to create. Additionally, personas may not be well suited to presenting detailed technical information, e.g., about disability, and their focus on representative individuals can make it more complex to capture the range of abilities in a population.

It is self-evident that storyboarding is not optimal for blind users, while it requires particular care for users with limited vision or color-blindness. On the contrary, it would appear to be a promising method for deaf or hearing-impaired users.

### 42.4.8 *Prototyping*

A prototype is a concrete representation of part or all of an interactive system. It is a tangible artifact, does not require much interpretation, and can be used by end users and other stakeholders to envision and reflect upon the final system (Beaudouin-Lafon and Mackay 2002).

Prototypes, also known as mockups, serve different purposes and thus take different forms:

- Off-line prototypes (also called paper prototypes) include paper sketches, illustrated story-boards, cardboard mock-ups and videos. They are created quickly, usually in the early stages of design, and they are usually thrown away when they have served their purpose.
- On-line prototypes, on the other hand, include computer animations, interactive video presentations, and applications developed with interface builders.

Prototypes also vary regarding their level of precision, interactivity and evolution. With respect to the latter, rapid prototypes are created for a specific purpose and then thrown away, iterative prototypes evolve, either to work out some details (increasing their precision) or to explore various alternatives, and evolutionary prototypes are designed to become part of the final system.

Research has indicated that the use of prototypes is more effective than other methods, such as interviews and focus groups, when designing innovative systems for people with disabilities, since potential users may have difficulty imagining how they might undertake familiar tasks in new contexts (Petrie et al., 1998). Using prototypes can be a useful starting point for speculative discussions, enabling the users to provide rich information on details and preferred solutions.

Prototypes are usually reviewed through user-trials, and therefore all considerations related to user trials and evaluation are pertinent. An obvious corollary is that prototypes must be accessible in order to be tested with disabled people. This may be easier to achieve with on-line prototypes, closely resembling the final system, than with paper prototypes.

### 42.4.9 *User trials*

In user trials, a product is tested by "real users" trying it out in a relatively controlled or experimental setting, following a standardized set of tasks to perform. User trials are performed for usability evaluation purposes. However, the evaluation of existing or competitive systems, or of early designs or prototypes, is also a way to gather user requirements (Maguire and Bevan 2002).

While there are wide variations in where and how a user trial is conducted, every user trial shares some characteristics. The primary goal is to improve the usability of a product by having participants who are representative of real users to use the product carrying out real tasks while being observed; the data that is collected is later analyzed. In field studies, the product or service is tested in a "real-life" setting.

In user trials, an appropriately equipped room needs to be available for each session. When planning the test, it should be taken into account that trials with elderly and users with disabilities may require more time than usual in order to complete the test without anxiety and frustration.

Research on the use of the most popular methods has indicated that modifications to well established user trial methods are necessary when users with disabilities are involved. For example, the think aloud protocol has been adapted to be applied differently when carrying out user trials with deaf users and blind users respectively (Chandrashekar et al., 2006; Roberts and Fels, 2006).

Furthermore, explicitly emphasizing during the instructions that it is the product that is being tested and not the user is very important (Poulson, Ashby and Richardson, 1996), since a trial may reveal serious problems with the product, to the extent that it may not prove possible to carry out some tasks. Therefore, it is important that users do not feel uncomfortable and attribute the product failure to their disability.

When the user trial participants are users with upper limb motor impairments and poor muscle control, it should be ensured that testing sessions are short, so as to prevent excessive fatigue.

Testing applications with children requires special planning and care. Guidelines have been developed to conduct usability testing with children. These guidelines provide a useful framework to obtain maximum feedback from children, while at the same time ensuring their comfort, safety and sense of well-being (Hanna, Risden and Alexander 1997). The need to involve children in every stage of design is particularly important in the case of children's technology, because for adult designers it is difficult (and often incorrect) to make assumptions about how a child may view or interpret data.

### 42.4.10 *Cooperative and Participatory Design*

Participatory design may adopt a wide variety of techniques, including brainstorming, scenario building, interviews, sketching, storyboarding and prototyping, with the full involvement of users.

Traditionally, partnership design techniques have been used for gathering user requirements from adult users. However, in the past few years a number of research projects have shown ways to adapt these techniques to benefit the design of technology process for non-traditional user groups, such as children and the elderly.

Cooperative inquiry has been widely used to enable young children to have a voice throughout the technology development process (Druin, 1999), based on the observation that although children are emerging as frequent and experienced users of technology, they were rarely involved in the development process. In these efforts, alterations were made to the traditional user requirement gathering techniques used in the process in order to meet the children's needs. For example, the adult researchers used note-taking forms, whereas the children used drawings with small amounts of text to create cartoon-like flow charts. Overall, involving children in the design process as equal partners was found to be a very rewarding experience and one that produced exciting results in the development of new technologies.

Designing technology applications to support older people in their homes has also shown an increase in necessity as the developed world is ageing. However, designing for this group of users is not an easy process as developers and designers often fail to fully grasp the problems that this user group faces when using technologies that affect their everyday life. HCI research methods need to be adjusted when used on this user group. They have to take into consideration that older adults experience a wide range of age-related impairments, including loss of vision, hearing, memory and mobility, which ultimately also contribute to loss of confidence and difficulties in orientation and absorption of information.

Participatory design techniques can help designers reduce the intergenerational gap between them and older people, and help better understand the needs of this group of users (Demirbileka and Demirkan, 2004). When older people participate in the design process from the start, their general fear towards using technology decreases, because they feel more in control and confident that the end result of the design process has truly taken into consideration their needs.

### 42.4.11 *Summary of design methods and techniques*

The table below (adapted from Antona et al., 2009), summarises the design methods and techniques discussed in the previous sections, suggesting an indicative path towards method selection for different target user groups.

| | Disability | | | | | |
|---|---|---|---|---|---|---|
| | Motion | Vision | Hearing | Cognitive / Communication | Children | Elderly |
| Direct observation | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Survey and questionnaires | ■ | ■ | ■ | ☒ | ■ | ■ |
| Interviews | ✓ | ✓ | ■ | ☒ | ■ | ■ |
| Activity diaries and cultural probes | ■ | ■ | ✓ | ■ | ■ | ✓ |
| Group discussions | ✓ | ✓ | ■ | ☒ | ■ | ■ |
| Empathic modeling | ✓ | ✓ | ✓ | ☒ | ☒ | ☒ |
| Scenarios, storyboards and personas | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Prototyping | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| User trials | ■ | ■ | ■ | ■ | ■ | ■ |
| Cooperative and participatory design | ✓ | ✓ | ✓ | ■ | ■ | ■ |

✓ Appropriate ■ Needs modifications and adjustments ☒ Not recommended

TABLE 42.1:

## 42.5 INTERACTION TECHNIQUES

### 42.5.1 *Speech*

Speech-based interactions allow users to communicate with computers or computer-related devices without the use of a keyboard, mouse, buttons, or other physical interaction devices. Speech-based interactions leverage a skill that is mastered early in life, and have the potential to be more natural than other technologies such as the keyboard. Speech-based interaction is of particular interest for children, older adults, and individuals with disabilities (Feng and Sears, 2009). Additionally, speech is a compelling input alternative when the user's hands are busy with another task (e.g., driving a car, conducting medical procedures) and the traditional keyboard and mouse may be inaccessible or inappropriate. Based on the input and output channels being employed, speech interactions can be categorized into three groups:

1. Speech output systems, which include applications that only utilize speech for output while leveraging other technologies, such as the keyboard and mouse, for input. Screen access software, which is often used by individuals with visual impairments, is an example of speech output.

2. Speech recognition systems, which include applications that utilize speech for input and other modalities for output, such as speech-based cursor control in a GUI (Graphical User Interface) and speech-based dictation systems.

3. Spoken dialogue systems, which include applications that utilize speech for both input and output, such as telephony systems and speech-based environment control systems with voice feedback.

Typical applications include:

- telephony systems which tend to use small input vocabularies as well as speech output, and environmental control applications with small input vocabularies which may support speech or non-speech output;

- speech-based interactions with graphical user interfaces which can support navigation, window manipulations, and various other command-based interactions with widely varying input vocabularies ranging from just a few words to several hundred;
- dictation applications which support users as they compose emails, letters, and reports as well as smaller tasks such as filling in portions of forms where free-form input is allowed.

Perhaps the most critical obstacle to date for the use of speech-based user interfaces has been recognition errors and the cumbersome recovery process.

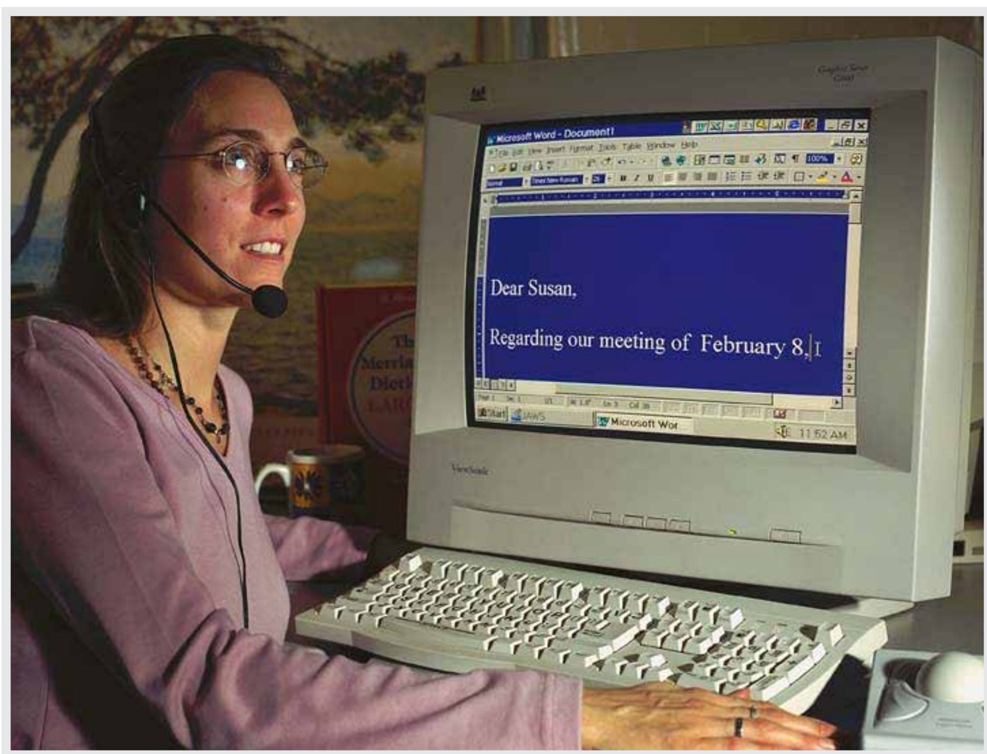





FIGURE 42.11: Speech interaction.

The use of small vocabularies can significantly reduce recognition error rates. Speech-based command and control systems allow users to interact by speaking predefined commands. Such systems often translate the user's utterances into system commands by matching the acoustic signal with models associated with each of the currently available commands.

Some systems use a simple menu, allowing users to select the appropriate option at each stage using simple spoken commands. Other systems try to establish a dialogue with the user in an attempt to gather the necessary information and complete the transaction as quickly as possible.

Multiple speech-based command and control solutions exist which emulate the conventional keyboard and mouse, allowing users to manipulate traditional graphical user interfaces via speech. These solutions are used by some individuals with physical disabilities who have difficulty using more traditional interaction solutions. Speech-based command and control systems are used for environmental control applications, allowing users to manipulate thermostats, lights, televisions, and any number of other devices.

Speech-based dictation systems can allow users to generate large quantities of text via speech. A speech recognizer is used to translate an audio signal into text. Dictation systems use a much larger vocabulary. As a consequence, dictation applications tend to be less accurate than command and control systems, and it is normally recommended that users create a personal speech profile to improve recognition accuracy. When feasible for the user, environment, and task, effective multimodal error correction solutions can provide an efficient alternative.

Large vocabulary dictation systems can provide a powerful alternative to the traditional keyboard and mouse, allowing the generation of a variety of documents such as emails, papers, and business reports. Importantly, the ability to produce such documents can significantly increase both educational and career opportunities. For the general public, such systems may serve as a useful alternative reducing the risk of keyboard or mouse based repetitive strain injuries.

### 42.5.2 *Haptics*

Hands have an impressive ability to perform different kinds of manipulation tasks, from working with miniature details to lifting heavy objects. Tactual information obtained via the skin has a basic importance for performance, but when the hands are functioning, there is a close cooperation between the sensors in the skin and sensors in the muscles, tendons and joints. The neural system coordinates efficiently the different kinds of sensory information with the muscles performing the movements.

Touch as a skin sense is often considered as a passive receiver of stimulation from the environment. However, the hand also relies on exploration to collect information. Active touch is often called *haptics*.

The use of haptics in interaction can support both vision and hearing, and offers additional possibilities to provide people with reduced functioning of senses such as vision and hearing with alternative options. The same interfaces may be used by all users in eyes-busy situations, or by visually impaired people in computer use. In addition, haptics also has the potential to help people with motor problems, as force feedback can be used to add extra strength or stability to motor actions (Jansson and Raisamo, 2009).

With respect to vision, sometimes haptics provides more advanced information, for instance concerning the weight of objects and the texture and hardness of surfaces. However, it lags behind vision in providing an overview of a scene. Obtaining an overview of a scene haptically may be a laborious and time-consuming task, requiring several explorations of the scene. Such differences are important to consider when dividing tasks between visual and haptic displays. The space covered by the two senses is also a very important difference. Vision allows the sensing of information kilometers away, while haptics is mainly restricted to the space within arm's reach.

One property of haptics is remote touching, that is, the experience of a distant object via some medium. The hand can pick up information via a tool and locate the information to the end of the tool. Visually impaired persons with a long cane can perceive the properties of the ground when touching it with the cane. Low-tech aids such as the long cane, Braille and embossed pictures are often used by the visually impaired. In all these cases, interaction between the haptic sense and the environment is important. Technically very simple aids, such as the long cane, can be very efficient if they make available the proper information and utilize the natural capacities of the haptic sense.

A visually impaired person can read text by using the hands to obtain a tactile equivalent of visual letters and other symbols, the most common version being Braille, coded within a six-point (or sometimes eight-point) rectangular matrix that is presented in embossed form and read by the fingers. There is also a mechanical Braille version consisting of matrices of pins. Matrices of point stimuli can be seen as representations of physical surfaces with low spatial resolution. Such matrices replace or supplement low-tech aids with technology providing more advanced information. One option is to present an extended matrix in contact with the skin. Such a matrix makes it possible to form a pattern by dynamically elevating some of the pins above the rest of the matrix. The pins can be either static or vibrating.



FIGURE 42.12: Braille display from www.handytech.de.

Haptic displays are for haptics what computer screens are for vision and loudspeakers for hearing. They are devices allowing the haptic manipulation of virtual objects and scenes via a computer. They are of interest both for people with vision as complements to visual and auditory information, and for people without vision as a substitute for this sense. Haptic displays provide force feedback by a "collision" between the endpoint of a stylus, a thimble or similar tool in a user's hand, and the surface of a virtual object. In addition to the shape of objects, surface properties such as hardness/softness, texture and friction can be rendered in 3D for haptic exploration.

The enormous potential of haptics when functioning in natural contexts is only partly utilized by the haptic displays developed so far. The most important constraints concern the number and size of contact surfaces. When the bare hand is used naturally, there are several contact surfaces and each of them has an extension of roughly at least a finger pad. In present day haptic displays, the number of contacts is quite low, in most cases just one. The contact surface is also, except in a few devices, only a tiny point. These differences with respect to natural haptics have important effects for the efficiency of haptic displays.



Figure 42.13: A haptic device.

### 42.5.3 *Scanning-based interaction*

Scanning is an interaction method addressing users with severe motor impairments in upper limbs (Ntoa et al, 2009). The main concept behind this technique is to eliminate the need for interacting with a computer application through traditional input devices, such as a mouse or a keyboard. Instead, users are able to interact with the use of switches. Scanning software makes the interactive objects composing a graphical user interface accessible through switches. It goes through the interactive interface elements and activates the element indicated by the user through pressing a switch. In most scanning software, interactive elements are sequentially focused and highlighted (e.g., by a colored marker). Furthermore, to eliminate the need for using a keyboard to type in text, an on-screen keyboard is usually provided.

During scanning, the focus marker scans the interface and highlights interactive objects sequentially, in a predefined order (e.g., from left to right and from top to bottom). Scanning can be either automatic or manual. In the first case, the marker automatically moves from one interface element to the next after a predefined time interval of user inactivity (i.e., not pressing the activation switch), while the time interval can usually be customized according to user needs. In manual scanning, the user moves the focus marker to the next interface element whenever she/he wishes with the use of a switch. Activation switches can vary from hand, finger, foot, tongue or head switches to breath-controlled switches or eye-tracking switches. Furthermore, any keyboard key (e.g., the space key) or mouse click can be used as a switch.

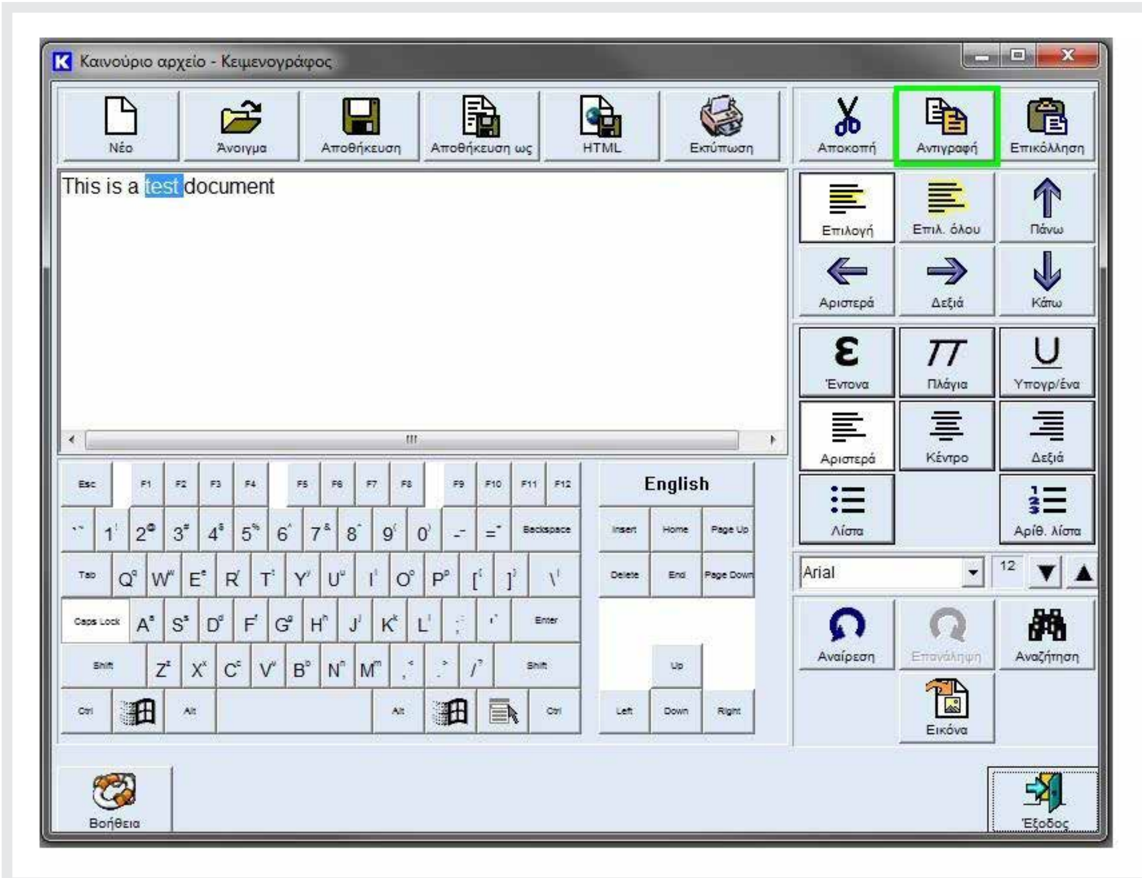





Figure 42.14: Scanning-based interaction.

There are several types of scanning techniques, mainly varying in their approach for accessing the individual interactive elements. The most popular scanning techniques include:

- Block scanning, in which items are grouped into blocks, aiming to minimize user input and enhance the interaction speed.
- Two-directional scanning, in which the user selects an element by

specifying its coordinates on the screen that is being scanned, first vertically, through a line that moves from the top of the screen towards the bottom, and then horizontally, through a pointer that moves along the selected horizontal line.

- Eight-directional scanning, which is used by several mouse emulation software products. In this method, the mouse pointer can be moved in one of eight directions, according to the user's preference. In order to achieve this, the pointer icon changes at specific time intervals to indicate one of the eight directions. The user selects the desired direction by pressing a switch and then the pointer starts moving in that direction. Once the pointer reaches the specific screen location that the user wishes to select, it can be stopped by a switch or key press.
- Hierarchical scanning, in which access to windows and window elements is provided according to their place in the window's hierarchical structure. Elements are usually divided into groups and subgroups according to their hierarchy (e.g., a toolbar acts as a container of the individual buttons it includes, a list box as a container of the included list items, etc.)
- Cluster scanning, in which elements on the screen are divided into clusters of targets, based on their locations.
- Adaptive scanning, in which the system's scan delay is adapted at runtime, based on measurements of user performance.

Applications with embedded scanning are developed so as to support scanning in the first place and are accessible to people with motor impairments. However, a user would need more than one application in order to carry out a variety of everyday computing tasks (e.g., web browser, email client, entertainment software, educational software, document authoring software, etc.). Scanning tools enable users to operate the graphical environment of the operating system, thus eliminating the need to use various specialized applications for carrying out everyday tasks. Keyboard and mouse emulation software with embedded scanning are popular among scanning tools, since they ensure user interaction without the need to use the traditional keyboard and mouse. A variety of approaches have been developed towards keyboard and mouse emulation (Evreinov, 2009). Mouse emulation with scanning support software allows users to control the mouse pointer using one or two switches.

### 42.5.4 *Eye tracking*

People use their eyes mainly for observation, but also use gaze to enhance communication. The direction of a person's gaze reveals their focus of visual attention, thus allowing eye gaze tracking for communication.

In some cases, eye gaze may be the only communication option available for an individual. For example, after a severe accident a person may not be able to speak, in which case, a doctor may ask the person to "look up" or "look down" as an indication of understanding and agreement. This method of communication can be expanded from a simple "yes" or "no" command to a communication system by adding meaningful objects in the view of a user. An example of this approach is the gaze communication board, where a board has pictures, commands or letters attached to it. The user selects items on the board by looking at them. The message is interpreted by other persons by following the user's eye. Such a system illustrates the simple communication power of eye gaze tracking.

Computer based gaze communication systems have been developed, where an eye tracking device and a computer replace the manual communication board (Majaranta, Bates and Donegan, 2009). In these eye tracking systems letters (or any other symbols, images or objects) are shown on a computer screen placed in front of the user. The user simply points and selects these items by looking at them, with an eye tracking device recording their eye movements and a computer program analyzing and interpreting their eye movements. Such a system forms a basic gaze communication system.

Recent advances in technology have considerably improved the quality of eye tracking systems, such that a far broader group of people may now benefit from eye control.

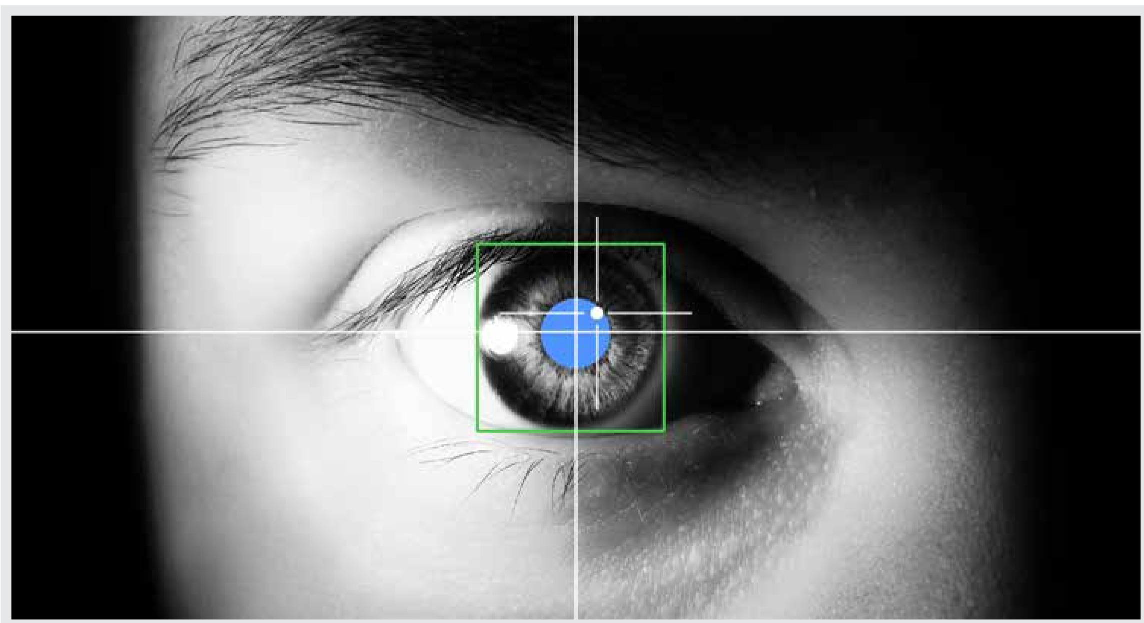




FIGURE 42.15: Eye tracking.

Normal eye movement is made from fixations on objects of interest (200 - 600 ms) joined by rapid saccades (30 - 120 ms) between those objects, with occasional smooth pursuit (follow a moving target) of moving objects. Since the fovea area of acute vision is fairly small, and people actually need to direct their gaze nearly directly towards an object of interest to get an acute view of the object (within 1 degree or so) this is what makes tracking the gaze direction possible: hence, if the eye is pointing at an object, the user is probably looking and perceiving that object. Eye movements are also largely unconscious and automatic; people do not normally need to think where to look. When needed, however, one can control gaze at will, making eye control possible.

The development of computing power has enabled gathering of eye tracking data in real time, as well as the development of assistive technology systems aimed directly at people with disabilities. Current eye tracking technologies have evolved into a range of technologies: electro-oculography (EOG) where the user wears small electrodes around the eye to detect the eye position; scleral contact lens/search coil where the user wears a contact lens with a magnetic coil on the eye that is tracked by an external magnetic system; video-oculography (VOG) or photo-oculography (POG) where still or moving images of the eye are taken to determine its position; and finally video-based combined pupil/corneal reflection techniques that extend VOG by artificially illuminating both the pupil and cornea of the eye for increased tracking accuracy. Most of the currently available eye control systems are video based with corneal reflection.

People who are unable to move but have good control over eye movements have traditionally been the best candidates for eye tracking systems. Immobility can make tracking easier in contrast to people who can move (voluntarily or involuntarily, for example in the case of cerebral palsy). However, eye control may still be a genuine choice for all users, as eye control can be faster and less tiring than, for example, a manual switch based system or a head pointing based system. Eye movements are extremely fast, and gaze pointing locates and points at a target long before a manually controlled mouse cursor may reach it.

Interpreting a person's intentions from their eye movements is not a trivial task. The eye is primarily a perceptual organ, not normally used for control, so the question arises of how casual viewing can be separated from intended gaze-driven commands. If all objects on the computer screen reacted to the user's gaze, this would cause the so-called “Midas touch” (or “Midas gaze”) problem: “everywhere you look something gets activated”. The obvious solution is to combine gaze pointing with some other modality for selection. If the person is able to produce a separate “click”, then this click can be used to select the focused item. This can be a switch, a blink, a wink, a wrinkle on the forehead or even smiling or any other available muscle activity. Blinks and winks can be detected from the same video signal used to analyze eye movements, removing the need for additional switch equipment.

If a user is only capable of moving their eyes, separate switches are not an option, and the system must be able to separate casual viewing from intentional eye control. The most common solution is to use dwell time, a prolonged gaze, with duration longer than a typical fixation (typically, 500-1000 ms). Most current eye control systems provide adjustable dwell time as one of the selection methods. Requiring the user to fixate for a long time does reduce false selections, but it is uncomfortable for the user, as fixations longer than 800 ms are often broken by blinks or saccades. Another solution for the Midas touch problem is to use a special selection area or an on-screen button.

Increasing the size of the targets on the screen improves the performance of eye gaze input. Making on-screen objects larger can make a difference between a user being able to use an eye-tracking device or not being able to use it at all. However, having only a few, large on-screen buttons at a time prevents the use of full-size keyboards such as a full “qwerty” keyboard. Instead, keys and controls can be organized hierarchically in menus and sub-menus, and special techniques such as automatic word prediction can be used to speed up the text entry process, with ambiguous or constantly changing and adapting keyboard layouts.

Gaze pointing, or placing the computer mouse cursor where the user is looking on the computer screen, is an intuitive method that requires little training, as it mimics the operation of a normal desktop mouse. Binding eye movements directly to mouse movements creates an “eye mouse”. This may seem an easy solution; however, there are several issues that have to be taken into account.

Eyes move constantly, and make small corrective movements even when fixating. If the cursor of such “eye mouse” followed eye movements without any smoothing, the cursor movement would appear very jerky and it would be difficult to concentrate on pointing, as the cursor itself would attract attention.

The main benefit of using mouse emulation is that it enables access to window-based graphical user interfaces. In addition, it enables the use of any existing access software, such as environmental control applications or “dwell click” tools. However, it should be noted that for a profoundly disabled person who does not have prior experience of any method of computer control, it may take time to master a gaze-pointing eye control system.

What makes certain eye tracking systems more suitable for people with disabilities are the applications (software) that are supported or come with the system.



FIGURE 42.16: Eye-tracking user interface.

Eye typing is typically the first application implemented and tried out by users with an eye control system. Eye typing can be slow, typically below 10 words per minute (wpm), due to dwell-time durations setting a limit on the maximum typing speed. When the user types by gaze, they cannot see the text appear in the text input field at the same time as they select a letter by "eye pressing" a key on an on-screen keyboard. To review the text written so far, the user needs to move their gaze from the on-screen keyboard to the typed text field. This shifting can be reduced by adding auditory feedback, e.g., an audible "click", or by speaking out each letter as they are written. Appropriate feedback also increases performance and improves accuracy. Providing appropriate feedback on the dwell-time progress and the selection process may significantly improve performance and make eye control more pleasant for the user.

When physically clicking a button, the user also feels and hears the button "click". Such extra confirming (auditory or tactile) feedback is missing when an "eye press" is used to click, and so must be provided. In addition to eye typing, there are several (dedicated) eye-controlled applications, such as eye drawing, eye music, internet browsing, email, games etc. Such applications are included in many of the commercial eye control systems targeted at people with disabilities.

Still, an extensive study on user requirements by Donegan et al. (2005) shows that, to date, eye control can effectively meet only a limited range of user requirements, and can only be used effectively by a limited number of people with disabilities. Furthermore, the range of applications that are suitable for easy and effortless control by the eye is limited.

### 42.5.5 *Gestures & head tracking*

The challenge of enriching a user's interaction with computer systems, beyond a typical mouse and a keyboard, was early recognized in the HCI field. Gestures are a powerful feature of human expression, either alone or as a means for augmenting spoken language.

Technology-wise, there are various approaches to gesture recognition, employing a variety of imaging and tracking devices or gadgets. Wearable devices are an unobtrusive solution to gesture-based interaction, functioning not only as an output but also as an input device. Gesture recognition is also possible using accelerometer-based information, available in numerous consumer electronics. Computer vision is also used for recognizing gestures through users' bare hands. Zabulis et al. (2009) identify a large number of methods proposed in the literature that utilize several types of visual features, such as skin color, shape, motion, and anatomical models of hands. Finally, recent research efforts have studied the issue of gesture recognition with the use of gaming devices such as Microsoft Kinect.



Figure 42.17: Wearable gesture recognition system.

Gesture-based interfaces can be employed in order to address the needs of people with disabilities, e.g. for home automation (Starner et al., 2000). However, a point of caution needs to be made over the interactions of people with disabilities and their accommodation by gesture-based systems, as there may be considerable limitations of the interaction for users with hand or cognitive disabilities. Also, care should be taken to provide the appropriate system feedback for gestural interactions of blind and low vision users. Gestural interfaces become even more challenging when addressing users with multiple limitations, due to age or illness.

Recent advances in technology allow exploiting not only hand gestures, but also the users' head and body movements as a means for providing information to computer systems. Computer control is a domain where head-based interaction is used, addressing mainly disabled or situationally disabled users. For example, head tracking using a stereo camera and a head-mounted tracking device has been used in cursor control and target selection tasks in desktop and interactive room environments (Morency et al., 2005). Alternative approaches suggest using the nose for cursor control and for hands-free games and interfaces in general (Gorodnichy and Roth, 2004), given that the nose, as the most protruding and the furthest from the axes of head rotation part of a person's face, has the largest degree of motion freedom. Another novel interaction technique employs a 3D audio radial pie menu that operated through head gestures for selecting items (Brewster et al., 2003). Users found such novel interactions comfortable and acceptable when tested under "eyes-free" mobile conditions. Commercial products towards computer control include the head mouse control, as well as head-controlled switches. Head mouse control devices use an infrared wireless optical sensor or a high resolution camera which tracks a tiny disposable target that is worn by the user in a convenient location on their forehead, glasses, hat, etc., thus translating the movements of the user's head into directly proportional movements of the computer mouse pointer.

### 42.5.6 *Brain Interfaces*

Brain Interfaces have been defined as a real-time communication system designed to allow a user to send messages voluntarily without sending them through the brain's normal output pathways such as speech, gestures or other motor functions, but only using bio-signals from the brain.

This type of communication system is needed by disabled individuals who have parts of their brain active, but have no means of communicating with the outside world. Brain Interfaces can provide new augmentative communications channels for those with severe motor impairments (Gnanayutham and George, 2009).

There are two types of Brain Interfaces, namely invasive (signals obtained by surgically inserting probes inside the brain), and non-invasive (electrodes placed externally on part of the body). The risks, difficulties and requirements involved in invasive Brain Interfaces make non-invasive Brain Interfaces the preferred choice for an assistive technology device. Non-invasive technology involves the collection of control signals for the Brain Interface without the use of any surgical techniques, with electrodes placed on the face, skull or other parts of the body. The signals obtained are first amplified, then filtered and thereafter converted from an analogue to a digital signal.

Bio potentials are electrical signals from the brain which can be obtained from skull, forehead or other parts of the body (the skull and forehead are predominantly used because of the richness of bio-potentials in these areas). Each bio-potential has its own unique characteristics, such as amplitude, frequency, method of extraction, and time of occurrence. Each brain-injured patient (apart from persistive vegetative state patients) can produce one or more of these bio-potentials with differing degrees of consistency. Brain injured patients can operate Brain Interfaces depending on the reliability of the bio-potential which they can produce. Current Brain Interfaces can transfer data up to 68 bits/second.

Electroencephalography measures electrical brain activity that results from thoughts or imagined movements. Electroencephalographic signals can be collected by electrodes placed on the scalp or forehead. Electromyographic and electrooculargraphic signals are the two front-runners for the most suitable bio-potentials for non-invasive Brain Interfaces. They are high amplitude bio-potentials, which can be more easily produced by a patient in comparison to other bio-potentials.

Various electrode positions are chosen by the developers, such as electrode caps, electrode headbands with different positions and number of electrodes or the International 10-20 System. The caps may contain as many as 256 electrodes, though typical caps use 16, 32, 64 or 128 positions, and each cap has its own potential sources of error. High-density caps can yield more information, but in practice they are hard to utilize for real time communications. This is due to the fact that the bio-potentials obtained from these large numbers of electrodes could possibly need extensive off-line processing to make any sense of what the user is trying to express. There is only one agreed standard for the positions and number of electrodes, which is the International 10-20 System of electrodes.

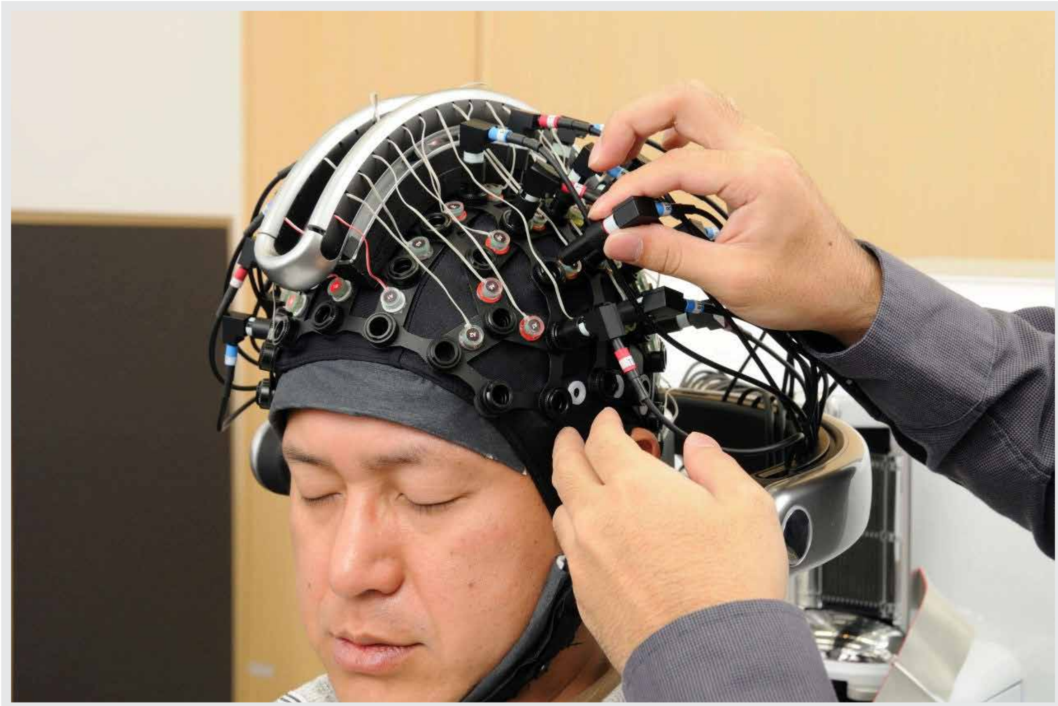




FIGURE 42.18: Brain Interface technology.

Brain interfaces can be used to communicate by using targets that contain words or phrases, switch devices on/off, launch computer applications, spelling by choosing individual words, dialing a telephone, surfing the net, operating rehabilitation robots, controlling wheel chairs etc.

Brain Interfaces so far have not been shown to be dependable enough for main software manufactures to integrate them into mainstream operating systems and applications. Many Brain Interface research applications are laboratory implementations, with limited test results obtained from the brain-injured community. The pace of research is increasing, and good progress is being made in the area of assistive technology.

This technology is promising but more evaluation has to be carried out with disabled participants in the field. Despite the potential shown by many Brain Interface devices, limited use of them is made by the disabled community. There is a clear need to take this technology outside the laboratory and into the field to nursing homes and hospitals. Research is also being conducted in wearable wireless Brain Interfaces where technology such as Bluetooth is proposed for transmitting and receiving signals from the participant.

### 42.5.7 *Sign language*

Interactions with computers involve reading or writing text. Deaf users, at first glance, would not appear to be disadvantaged in their ability to read and write. However, interfaces requiring reading and writing also have the potential to disenfranchise many deaf users.

Millions of deaf and hard of hearing people worldwide use a sign language to communicate. Sign languages are naturally occurring languages with linguistic structures (e.g., grammars, vocabularies, word order, etc.) distinct from spoken languages. For instance, American Sign Language (ASL) is the primary means of communication for an estimated half a million people in the United States. ASL is a full natural language that includes various linguistic phenomena that make it distinct from English. Sign languages are not based on the spoken languages of the region. ASL is a visual language in which the signer's facial expression, eye gaze, head movement, shoulder tilt, arm movements, and hand shapes convey linguistic information; however, it is not enough to know how a signer's body moves in order to understand an ASL sentence. It is also necessary to remember how the "signing space" around their body has been filled with imaginary placeholders that represent the entities under discussion.

There are a number of factors that determine whether an individual with hearing loss will use a sign language, including their family circumstances, educational experience, age of onset of hearing loss, and degree of hearing loss. Signers comprise a Deaf Community, whose membership is determined more by a shared language rather than by degree of hearing loss.

In fact, people who experience hearing loss as adults tend not to become signers or members of this community. Contrary to popular expectation, sign languages are not universal; Deaf Communities around the world have their own native sign languages. Deaf individuals often acquire a sign language as their first language and are most fluent and comfortable in this first language.

Few computer user interfaces make sufficient accommodation for deaf users. Despite the fact that many deaf individuals are skilled readers, not all deaf signers develop this level of proficiency. For example, studies have shown that the majority of deaf high school graduates in the United States have only a fourth grade English reading level—this means that deaf students around age 18 have a reading level more typical of 10-year-old hearing students. Sign language interfaces are a necessity for that subset of the deaf population with difficulty in reading and writing (Huenerfauth and Hanson, 2009).

A machine translation system from English text into ASL animations could increase the accessibility of user interfaces for signers. Instead of presenting written text on a television screen, telephone display, or computer monitor, each could instead display ASL signing.

In addition, technologies for recognizing sign language could also benefit deaf signers. The ability to input commands to a computing system using ASL would make the interaction more natural for deaf signers, and the ability of the system to translate sign language input into English text or speech could open additional avenues of communication for deaf signers with low levels of English literacy.

The ultimate sign language interface tool would be one that could recognize sign language input while also having the ability to output sign language from spoken utterances or text. Such a tool would allow easy interaction between deaf signers and hearing speakers. It would also allow deaf signers natural and easy access to computers and other devices. However, today, both production and, even more so, recognition systems are in relatively early stages of development.

A number of applications have been developed that display videos of humans performing sign language. These interfaces have been employed not only for making audio and speech materials accessible to signers, but also for teaching reading and writing to deaf signers. While using videos of human sign language performances can be appropriate when there are a finite set of sentences that a system must convey to the user, it is difficult to use videos as the basis for a computer system that must generate/assemble novel signed sentences.

Most successful sign language generation systems have instead chosen to create animations of a 3D human-like character that moves to perform a sign language message. This approach has the advantage of allowing the system to blend more easily together individual signs into a smooth-looking sign language sentence. Research into virtual reality, human modeling and animation has reached a point of sophistication where it is now possible to construct a human model which is articulate and responsive enough to perform sign languages. The level of quality of such human avatar animations has increased such that human signers can now view the onscreen animations and successfully interpret the movements of the avatar to understand its meaning.



Figure 42.19: A signing avatar.

Sign Language Translation Systems analyze the linguistic structure of the input text. The grammatical structure, word order, and vocabulary of the text are translated into the appropriate sign language grammatical structure, word order, and vocabulary. Such systems produce a script that specifies the sign language performance—generally using sign language synthesis software to produce an actual animation output in which a human-like character performs the sign language sentence.

Sign recognition has the goal of automatically converting the sign language performance of a human user into a computational representation of the performance; this allows the computer to identify the meaning of the user's signing and possibly to later translate it into text or speech.

Educational software can be created to help users learn sign language literacy skills (by watching sign language animations or performing sign language that is recognized by the system) or to help users learn other academic content (through explanation in the form of sign language animation). Sign language scripting software can also be created to allow users to create and edit sign language animations much as word processing software allows editing of written language content.

In the near future, deaf signers may be able to benefit from machine sign language translation technologies in various applications as long as they are aware that signed translations may not be as accurate as those provided by human interpreters. Given the relatively young state-of-the-art of sign language technologies, service providers (e.g., governments, companies, media outlets, etc.) must be careful not to deploy these technologies prematurely in the name of accessibility.

### 42.5.8 *Multimodal interfaces*

Multimodal interaction is a characteristic of everyday human discourse. People speak, shift eye gaze, gesture, and move in an effective flow of communication.

Multimodal user interfaces enrich interaction with these elements of natural human behavior. Multimodal systems process two or more combined user input modes—such as speech, pen, touch, manual gestures, gaze and head and body movements—in a coordinated manner with multimedia system output (Oviatt, 2003). They are "multi-sensory" (i.e., they utilize multiple sensory modalities), "multi-channel" (i.e., they utilize multiple channels, on the same or different modalities), "multi-tasking" (i.e., they allow users to perform several tasks at the same time), and "multi-form" (i.e., they allow users perform the same tasks in alternative ways).

Research on modalities and multimodality in human-computer interaction has had major beneficial effects for Design for All. The increase in the range of available modalities and styles of multimodal interaction makes it possible to compensate for a growing diversity of physical disabilities, and thus to provide a larger community of disabled users with easier computer access and appropriate facilities for browsing and processing digital information, reducing the need for assistive technologies. Multimodal interfaces have the potential to greatly expand the accessibility of computing to diverse non-specialist users and they will "increase the accessibility of computing for users of different ages, skill levels, cognitive styles, sensory and motor impairments, native languages, or even temporary illnesses" (Oviatt and Cohen, 2000).

For example, multimodal interfaces have features that could help to overcome some of the limitations of existing solutions for the blind. In particular, they could provide better adaptation to users' needs, more intuitive, concise and quick interaction modes, easier learning and reduction of memorization efforts, and could increase power expression (Bellik and Burger, 1994). Different modalities can be used concurrently, so as to increase the quantity of information made available or present the same information in different contexts, or redundantly, to address different interaction channels, both to reinforce a particular piece of information or to cater for the different abilities of users (Antona et al., 2007).

Recent advances in the processing of novel input modalities, such as speech, gestures, gaze, or haptics and synergistic combinations of modalities, provide appropriate substitutes for direct manipulation in situations where the use of a keyboard, mouse, and standard screen is awkward or impossible (Carbonell, 2009).

### 42.5.9 *Summary of Interaction Techniques*

The table below summarises the interaction techniques discussed in the previous sections, the related target user groups and the most typical applications.

| | Target Groups | Typical Applications |
|---|---|---|
| Speech | Visually impaired<br>Motor Impaired*<br>Children*<br>Elderly* | Telephony systems<br>Speech-based interaction<br>Dictation systems |
| Haptics | Visually impaired<br>Hearing Impaired<br>Motor Impaired* | Braille displays<br>Remote sensing<br>Haptic displays |
| Scanning-based Interaction | Motor / speech Impaired<br>Cognitively impaired | Applications with embedded scanning<br>Scanning tools<br>Keyboard and mouse emulation |
| Eye-tracking | Motor Impaired | Gaze communication systems<br>Mouse emulation<br>Eye-typing |

| | | |
|---|---|---|
| Gesture and head tracking | Motor / speech Impaired*<br>Visually impaired<br>Hearing Impaired<br>Children Elderly | Home automation<br>Mouse emulation<br>Target selection applications<br>Head-controlled switches |
| Brain Interfaces | Motor / speech Impaired | Communication systems<br>Computer control<br>Wheelchair control |
| Sign Language | Hearing Impaired | Sign recognition<br>Sign language generation |

** with the necessary provisions / modifications*

## 42.6 FUTURE DIRECTIONS

As a result of the increasing demand for ubiquitous and continuous access to information and services, interactive technologies are evolving towards a new interaction paradigm referred to as Ambient Intelligence (AmI).

***Definition***

***Ambient Intelligence:*** *In an AmI world, massively distributed devices operate collectively while embedded in the environment using information and intelligence that is hidden in the interconnection network. Lighting, sound, vision, domestic appliances, personal health-care devices, and distributed services all cooperate seamlessly with one another to improve the total user experience through the support of natural and intuitive user interfaces. In short, Ambient Intelligence refers to electronic systems that are sensitive and responsive to the presence of people.*

Aarts & de Ruyter, 2009

Ambient Intelligence will have profound consequences on the type, content and functionality of the emerging products and services, as well as on the way people will interact with them, bringing about multiple new requirements. The potential of AmI environments to address older and disabled people's everyday life needs is expected to have a radical impact on social inclusion and independent living. Many applications and services which address a wide variety of issues critical for older and disabled people are already becoming available, for example in the domain of Ambient Assisted Living (AAL), and are targeted to make possible and enjoyable a more independent, active and healthy life. A number of ICT solutions address daily and independent living in areas such as social communication, daily shopping, travel, social life, public services, safety, reminders, telecare and telemedicine, personal health systems, and support for people with cognitive problems and their carers (The European strategy in ICT for Ageing Well of 2010).

User-friendly interfaces are necessary for all sorts of equipment in the home and outside, taking into account that many older people have impairments in vision, hearing, mobility or dexterity. Clearly, the benefits of AmI environments can only be fully achieved and accepted by their target end-users if such technologies can demonstrably be developed in such a way as to guarantee inclusive accessibility for a wide variety of functional limitations brought about by age or disabilities.

In such a dynamically evolving and complex technological environment, accessibility and usability by users with different characteristics and requirements cannot be addressed through solutions introduced once the main building components of the new environment are in place. In such a context, the concept of Design for All acquires critical importance towards streamlining accessibility into the new technological environment through generic solutions (Emiliani and Stephanidis, 2005). However, in the context of Ambient Intelligence, Design for

All will need to evolve in order to address a series of new challenges posed by the evolving technological environment.

The accessibility of AmI environments poses different problems and is more complex than currently available approaches to the accessibility of desktop or web applications and services, as AmI environments do not simply introduce a new technology, but an integrated set of technologies. Different levels of accessibility may be distinguished. A first level concerns accessibility of individual devices. Interactive devices need to be accessible to their owners according to their needs, but basic accessibility should also be provided for other users with potentially different needs. A second level concerns the accessibility of the environment as a whole, intended as equivalent access to content and functions for users with diverse characteristics, not necessarily through the same devices, but through a set of dynamic interaction options integrated in the environment.

It is likely that some of the built-in features of AmI environments, such as multi-modality, will facilitate the provision of solutions that will be accessible by design. For example, blind users will benefit from the wider availability of voice input and output. A novel aspect is that in AmI environments, the accessibility of the physical and of the virtual world need to be combined. For example, for blind, visually impaired and motor-impaired users, requirements related to interaction need to be combined with requirements related to physical navigation in the interactive environment.

As a result, developing truly accessible AmI environments is currently very expensive in terms of time, effort, costs and required knowledge, and the results are often of limited flexibility and reusability in terms of accessibility solutions and target user groups addressed.

Therefore, it is necessary to elaborate on the one hand a user-centered but also context-aware methodology for enabling Design for All in Ambient

Intelligence, and on the other hand modern tools and personalized assistive solutions that will constitute the building blocks for the development of independent living AmI environments addressing the interaction needs of older and disabled persons.

Several challenges need to be addressed in order to elaborate a systematic approach to accessibility and Design for All in AmI environments (Margetis et al., 2012):

- Advancing knowledge of user requirements and of the appropriateness of different solutions for different combinations of user characteristics / functional limitations and environment characteristics / functions, and creating related ontological models.
- Developing reference architectural models that will accommodate system requirements inherent in the Design for All needs in AmI environments, while allowing for accessible multi-modal interaction.
- Providing ready-to-use accessibility solutions supporting alternative interaction techniques for various combinations of user abilities / functional limitations.
- Developing design tools for accessible AmI environments.
- Developing accessible AmI applications in key everyday life domains, such as, home, work, learning, health and self-care.
- Evaluating the developed assistive solutions tools and applications in order to assess their accessibility, usability and added value for the target users.

## 42.7 WHERE TO LEARN MORE

### 42.7.1 *Books*

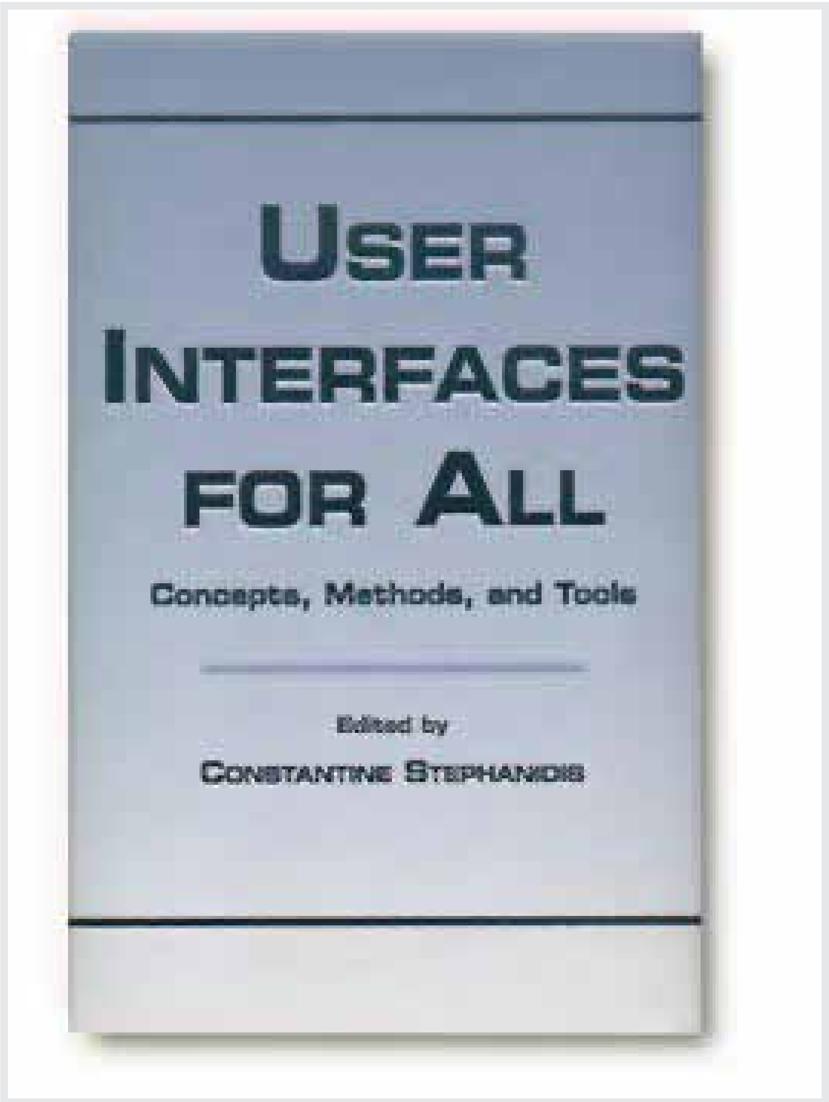





Figure 42.20: USER INTERFACES FOR ALL: Concepts, Methods, and Tools. Editor: Constantine Stephanidis. Publisher:Lawrence Erlbaum Associates. ISBN:0-8058-2967-9. More Info.

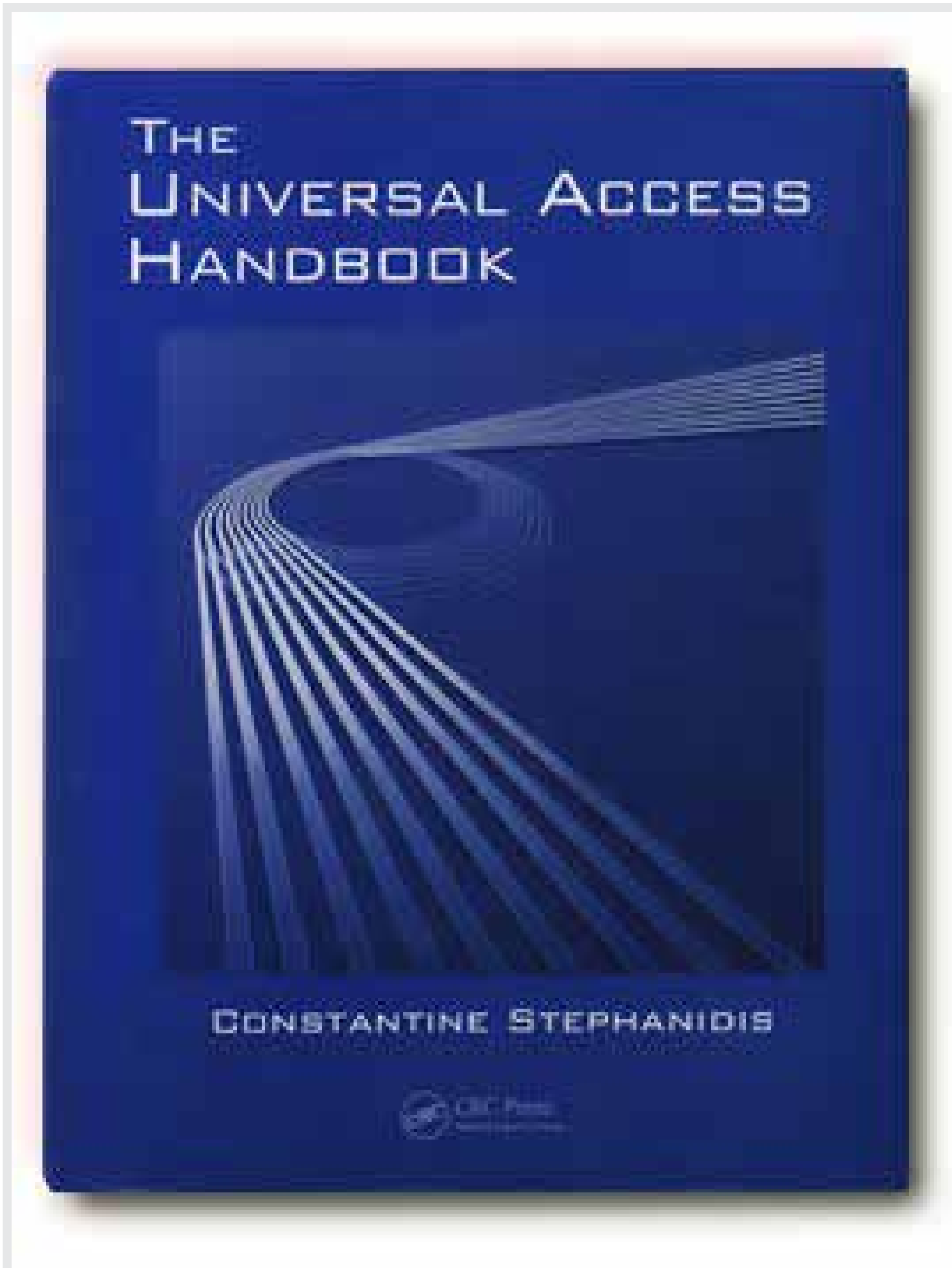
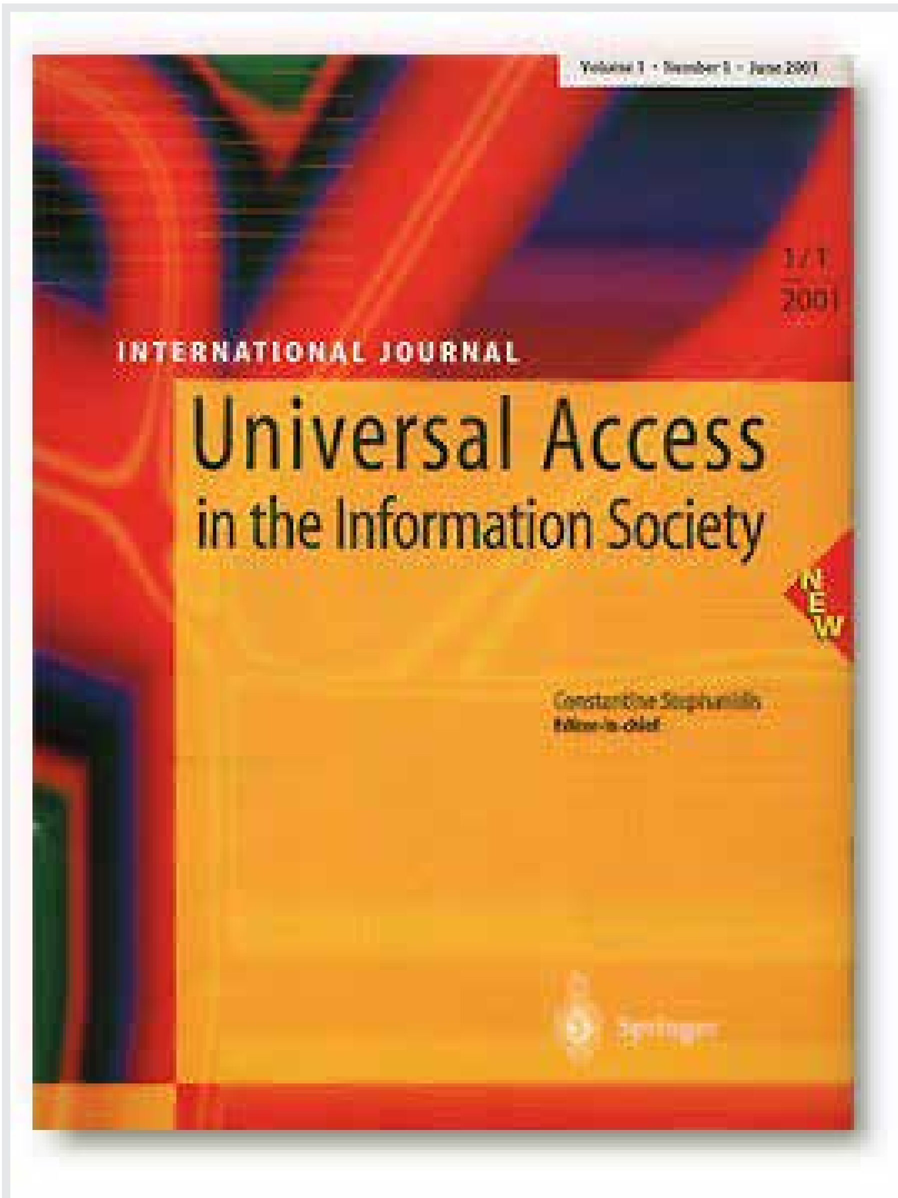





FIGURE 42.21: THE UNIVERSAL ACCESS HANDBOOK. Editor: Constantine Stephanidis. Publisher: CRC Press Taylor & Francis Group. ISBN: 978-0-8058-6280-5. More Info.

## 42.7.2 *Journal*




*Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms below.*

Figure 42.22: Universal Access in the Information Society. Editor-in-chief: Constantine Stephanidis. Publisher: Springer-Verlag Heidelberg. More Info.

### 42.7.3 *Conference*

Universal Access in Human-Computer Interaction, affiliated with the Human-Computer Interaction International Conference Series

## 42.8 REFERENCES


*Aarts, Emile and Ruyter, Boris de (2009): New research perspectives on Ambient Intelligence. In Journal of Ambient Intelligence and Smart Environments, 1 (1) pp. 5-14*

*Adobe Systems Incorporated (2014). Adobe accessibilitÃ . Retrieved 23 January 2014 from Adobe Systems Incorporated: http://www.adobe.com/accessibility.html*

*Aging, Administration on (2009). A Profile of Older Americans: 2009. US Department of Health and Human Services http://www.aoa.gov/aoaroot/aging_statistics/profile/2009/docs/2009profile_508.pdf*

*Antona, Margherita, Ntoa, Stavroula, Adami, Ilia and Stephanidis, Constantine (2009): User Requirements Elicitation for Universal Access. In: Stephanidis, Constantine (ed.). "The Universal Access Handbook". Boca Ration, USA: Taylor and Francispp. 15-1-15-14*

*Antona, Margherita, Burzagli, Laura, Emiliani, Pier L. and Stephanidis, Constantine (2007): The ISTAG scenarios: a case study. In: Roe, Patrick R.W. (ed.). "Towards an Inclusive Future: Impact and Wider Potential of Information and Communication Technologies". Brussels, Belgium: COST219terpp. 158-187*

*Asakawa, Chieko and Leporini, Barbara (2009): Screen Readers. In: Stephanidis, Constantine (ed.). "The Universal Access Handbook". Boca Ration, USA: Taylor and Francispp. 28-1-28-18*

*Ashok, Mahima and Jacko, Julie A. (2009): Dimensions of user diversity. In: Stephanidis, Constantine (ed.). "The Universal Access Handbook". Boca Ration, USA: Taylor and Francispp. 4-1-4-16*

*Assistive Technology Act of 1998. http://www.section508.gov/docs/AssistiveTechnologyActOf1998Full.pdf*

*Basdekis, Ioannis, Klironomos, Iosif, Metaxas, Ioannis and Stephanidis, Constantine (2010): An overview of web accessibility in Greece: a comparative study 2004--2008. In Universal Access in the Information Society, 9 (2) pp. 185-190*

*Beaudouin-Lafon, Michel and Mackay, Wendy (2007): Prototyping Development and Tools.*

*In: Sears, Andrew and Jacko, Julie A. (eds.). "The human-computer interaction handbook: fundamentals, evolving technologies and emerging applications book contents". Mahwah, USA: Lawrence Erlbaum Associatespp. 1006-1031*

*Bellik, Yacine and Burger, Dominique (1994): Multimodal Interfaces: New Solutions to the Problem of Computer Accessibility for the Blind. In: Adelson, Beth, Dumais, Susan and Olson, Judith S. (eds.) Proceedings of the Human Factors in Computing Systems Conference, CHI 94 April 24-28, 1994, Boston, Massachusetts. pp. 267-268*

*Beyer, Hugh R. and Holtzblatt, Karen (1997): Contextual Design: Defining Customer-Centered Systems (Interactive Technologies). Morgan Kaufmann*

*Brewster, Stephen A., Lumsden, Joanna, Bell, Marek, Hall, Malcolm and Tasker, Stuart (2003): Multimodal 'eyes-free' interaction techniques for wearable devices. In: Cockton, Gilbert and Korhonen, Panu (eds.) Proceedings of the ACM CHI 2003 Human Factors in Computing Systems Conference April 5-10, 2003, Ft. Lauderdale, Florida, USA. pp. 473-480*

*Bruckman, Amy and Bandlow, Alisa (2002): Human-computer interaction for kids. In: Jacko, Julie A. and Sears, Andrew (eds.). "The human-computer interaction handbook". Hillsdale, USA: L. Erlbaum Associatespp. 428-440*

*Bruseberg, Anne and Mcdonagh-Philp, Deana (2001): New product development by eliciting user experience and aspirations. In International Journal of Human-Computer Studies, 55 (4) pp. 435-452*

*Carbonell, NoÃ«lle (2009): Contributions of "Ambient" Multimodality to Universal Access. In: Stephanidis, Constantine (ed.). "The Universal Access Handbook". Boca Ration, USA: Taylor and Francispp. 40-1-40-24*

*Carroll, John M. (ed.) (1995): Scenario-Based Design: Envisioning Work and Technology in System Development. John Wiley and Sons*

*Castells, Manuel (1999). Information Technology, Globalization and Social Development. Retrieved 23 January 2014 from United Nations Research Institute for Social Development: http://www.unrisd.org/80256B3C005BCCF9/%28httpAuxPages%29/F270E0C066F3DE7780256B67005B728C/$file/dp114.pdf*

*Chandrashekar, Sambhavi, Stockman, Tony, Fels, Deborah and Benedyk, Rachel (2006): Using think aloud protocol with blind users: a case for inclusive usability evaluation methods. In: Eighth Annual ACM Conference on Assistive Technologies 2006. pp. 251-252*

*Commission, European (2010). The European strategy in ICT for Ageing Well. Retrieved 25 January 2014 from European Commission: http://ec.europa.eu/digital-agenda/life-and-work*

Cooper, Alan (1999): The Inmates Are Running the Asylum: Why High Tech Products Drive Us Crazy and How to Restore the Sanity. Sams

Crabtree, Andy, Hemmings, Terry, Rodden, Tom, Cheverst, Keith, Clarke, Karen and Dewsbury, Guy (2003): Designing with Care: Adapting Cultural Probes to Inform Design in Sensitive Settings. In: Viller, Stephen and Wyeth, Peta (eds.) Proceedings of OzCHI 2003, New Directions in Interaction Information environments, Media and Technology November 26-28, 2003, Brisbane, Australia. pp. 4-13

Demirbilek, Oya and Demirkan, Halime (2004): Universal product design involving elderly users: a participatory design model. In Applied Ergonomics, 35 (4) pp. 361-370

Dickinson, Anna, Eisma, Roos, Syme, Audrey and Gregor, Peter (2002): UTOPIA: Usable Technology for Older People: Inclusive and Appropriate. In: Brewster, Stephen and Zajicek, Mary (eds.) Proceedings of BCS HCI, A New Research Agenda for Older Adults September 2-6, 2002, London, United Kingdom. pp. 38-39

Donegan, Mick, Oosthuizen, Lisa, Bates, Richard, Daunys, Gintautas, Hansen, John P., Joos, Markus, Majaranta, PÃ¤ivi and Signorile, Isabella (2005). D3. 1 User requirements report with observations of difficulties users are experiencing. Retrieved 23 January 2014 from COGAIN: http://wiki.cogain.org/images/e/ef/COGAIN-D3.1.pdf

Doulgeraki, Constantina, Partarakis, Nikolaos, Stephanidis, Constantine and Mourouzis, Alexandros (2009): Methods and Tools for the Development of Unified Web-Based User Interfaces. In: Stephanidis, Constantine (ed.). "The Universal Access Handbook". Boca Ration, USA: Taylor and Francispp. 23-1-23-26

Druin, Allison (1999): Cooperative Inquiry: Developing New Technologies for Children with Children. In: Altom, Mark W. and Williams, Marian G. (eds.) Proceedings of the ACM CHI 99 Human Factors in Computing Systems Conference May 15-20, 1999, Pittsburgh, Pennsylvania. pp. 592-599

Eisma, R., Dickinson, A., Goodman, J., Syme, A., Tiwari, L. and Newell, Alan F. (2004): Early user involvement in the development of information technology-related products for older people. In Universal Access in the Information Society, 3 (2) pp. 131-140

Engelen, Jan (2009): eAccessibility Standardization. In: Stephanidis, Constantine (ed.). "The Universal Access Handbook". Boca Ration, USA: Taylor and Francispp. 55-1-55-9

Evreinov, Grigori (2009): Virtual Mouse and Keyboards for Text Entry. In: Stephanidis, Constantine (ed.). "The Universal Access Handbook". Boca Ration, USA: Taylor and Francispp. 29-1-29-22

*Feng, Jinjuan and Sears, Andrew (2009): Speech Input to Support Universal Access. In: Stephanidis, Constantine (ed.). "The Universal Access Handbook". Boca Ration, USA: Taylor and Francispp. 30-1-30-16*

*Gaver, Bill, Dunne, Tony and Pacenti, Elena (1999): Design: Cultural probes. In Interactions, 6 (1) pp. 21-29*

*Gnanayutham, Paul and George, Jennifer (2009): Brain-Body Interfaces. In: Stephanidis, Constantine (ed.). "The Universal Access Handbook". Boca Ration, USA: Taylor and Francispp. 37-1-37-12*

*Gorodnichy, Dmitry O. and Roth, Gerhard (2004): Nouse 'use your nose as a mouse' perceptual vision technology for hands-free games and interfaces. In Image and Vision Computing, 22 (12) pp. 931-942*

*Grammenos, Dimitrios, Savidis, Anthony and Stephanidis, Constantine (2005): UA-Chess: A Universally Accessible Board Game. In: Stephanidis, Constantine (ed.) Proceedings of the 11th International Conference on Human-Computer Interaction July 22-27, 2005, Las Vegas, USA.*

*Hall, Lynne and Mallalieu, Gill (2003): Identifying the Needs and Expectations of Users with Learning Disabilities. In: Stephanidis, Constantine (ed.) Proceedings of the Tenth International Conference on Human-Computer Interaction June 22-27, 2003, Crete, Greece. pp. 837-841*

*Hanna, Libby, Risden, Kirsten and Alexander, Kirsten (1997): Guidelines for Usability Testing with Children. In Interactions, 4 (5) pp. 9-14*

*Huenerfauth, Matt and Hanson, Vicki L. (2009): Sign Language in the Interface: Access for Deaf Signers. In: Stephanidis, Constantine (ed.). "The Universal Access Handbook". Boca Ration, USA: Taylor and Francispp. 38-1-38-18*

*IBM (2014). IBM Accessibility. Retrieved 24 January 2014 from IBM: http://www-03.ibm.com/able/*

*IMS Global Learning Consortium (2002). IMS Guidelines for Developing Accessible Learning Applications. Retrieved 24 January 2014 from IMS Global Learning Consortium: http://www.imsglobal.org/accessibility/accessiblevers/index.html*

*ISO 9241-210 (2010) Ergonomics of human-system interaction -- Part 210: Human-centred design for interactive systems*

*Keates, Simeon (2009): Motor Impairments and Universal Access. In: Stephanidis, Constantine* *(ed.). "The Universal Access Handbook". Boca Ration, USA: Taylor and Francispp. 5-1-5-14*

Kemppainen, Erkki, Kemp, John D. and Yamada, Hajime (2009): *Policy and Legislation as a Framework of Accessibility. In:* Stephanidis, Constantine *(ed.). "The Universal Access Handbook". Boca Ration, USA: Taylor and Francispp. 53-1-53-16*

Kurniawan, Sri *(2009): Age-Related Differences in the Interface Design Process. In:* Stephanidis, Constantine *(ed.). "The Universal Access Handbook". Boca Ration, USA: Taylor and Francispp. 8-1-8-12*

Lewis, Clayton *(2009): Cognitive Disabilities. In:* Stephanidis, Constantine *(ed.). "The Universal Access Handbook". Boca Ration, USA: Taylor and Francispp. 7-1-7-12*

Macaulay, Linda A. *(1996): Requirements engineering. Berlin, Germany,*

Mace, Ronald L., Hardie, Graeme J. and Place, Jaine P. *(1991): Accessible environments: Toward universal design. In:* Preiser, Wolfgang, Vischer, Jacqueline C. and White, Edward T. *(eds.). "Design interventions: Toward a more human architecture". New York, USA: Van Nostrand Reinholdp. 155 -176*

Maguire, Martin and Bevan, Nigel *(2002): User Requirements Analysis: A Review of Supporting Methods. In: Proceedings of the IFIP 17th World Computer Congress - TC13 Stream on Usability Gaining a Competitive Edge August 25-30, 2002, Montreal, Canada. pp. 133-148*

Majaranta, Paivi, Donegan, Mick and Bates, Richard *(2009): Eye Tracking. In:* Stephanidis, Constantine *(ed.). "The Universal Access Handbook". Boca Ration, USA: Taylor and Francispp. 36-1-36-19*

Margetis, George, Antona, Margherita, Ntoa, Stavroula and Stephanidis, Constantine *(2012): Towards accessibility in ambient intelligence environments. In:* Paterno, Fabio, Ruyter, Boris de, Markopoulos, Panos, Santoro, Carmen, Loenen, Evert van and Luyten, Kris *(eds.) Proceedings of the 3rd International Joint Conference in Ambient Intelligence November 13-15, 2012, Pisa, Italy. pp. 328-337*

*Microsoft (2013). Designing Accessible Applications. Retrieved 24 January 2014 from Microsoft: http://msdn.microsoft.com/en-us/library/aa291864%28v=vs.71%29.aspx*

Morency, Louis-Philippe, Sidner, Candace L., Lee, Christopher and Darrell, Trevor *(2005): Contextual recognition of head gestures. In:* Lazzari, Gianni, Pianesi, Fabio, Crowley, James L., Mase, Kenji and Oviatt, Sharon L. *(eds.) Proceedings of the 7th International Conference on Multimodal Interfaces - ICMI 2005 October 4-6, 2005, Trento, Italy. pp. 18-24*

Nees, Michael A. and Walker, Bruce N. *(2009): Auditory Interfaces and Sonification. In:*

*Stephanidis, Constantine (ed.). "The Universal Access Handbook". Boca Ration, USA: Taylor and Francispp. 33-1-33-16*

*Newell, Alan F., Arnott, John L., Carmichael, Alex and Morgan, Maggie (2007): Methodologies for Involving Older Adults in the Design Process. In: Stephanidis, Constantine (ed.) UAHCI 2007 - 4th International Conference on Universal Access in Human-Computer Interaction - Part 1 July 22-27, 2007, Beijing, China. pp. 982-989*

*Nicolle, Colette and Maguire, Martin C. (2003): Empathic Modelling in Teaching Design for All. In: Stephanidis, Constantine (ed.) Proceedings of the Tenth International Conference on Human-Computer Interaction June 22-27, 2003, Crete, Greece. pp. 143-147*

*Ntoa, Stavroula, Savidis, AnthÎ¿ny and Stephanidis, CÎ¿nstantine (2009): Automatic Hierarchical Scanning for Windows Applications. In: Stephanidis, Constantine (ed.). "The Universal Access Handbook". Boca Ration, USA: Taylor and Francispp. 35-1-35-16*

*Oppenheim, Bram (2000): Questionnaire Design. Continuum International Publishing Group*

*Oviatt, Sharon L. (2003): Multimodal interfaces. In: Jacko, Julie A. and Sears, Andrew (eds.). "The human-computer interaction handbook: fundamentals, evolving technologies and emerging applications". Lawrence Erlbaum Associatespp. 286-304*

*Oviatt, Sharon L. and Cohen, Philip R. (2000): Multimodal Interfaces That Process What Comes Naturally. In Communications of the ACM, 43 (3) pp. 45-53*

*Paramythis, Alexandros, Stephanidis, Constantine and Savidis, Anthony (2005): Developing Adaptive Interfaces for the Web. In: Vu, Kim-Phuong L. and Proctor, Robert W. (eds.). "Handbook of Human Factors in Web Design". Mahwah, USA: Lawrence Erlbaum Associatespp. 251-266*

*Pernice, Kara and Nielsen, Jakob (2001). Beyond ALT Text: Making the Web Easy to Use for Users with Disabilities. Nielsen Norman Group http://media.nngroup.com/media/reports/free/Usability_Guidelines_for_Accesible_Web_Design.pdf*

*Petrie, Helen, Johnson, Valerie, Furner, Stephen and Strothotte, Thomas (1998): Design Lifecycles and Wearable Computers for Users with Disabilities. In: Johnson, Chris (ed.) Proceedings of the First Workshop on Human Computer Interaction with Mobile Devices May 21-23, 1998, Glasgow, Scotland.*

*Poulson, David, Ashby, Martin and Richardson, Simon (1996): USERfit-A Practical Handbook on User-Centred Design for Assistive Technology. Brussels, Belgium, European Commission*

*Raisamo, Roope and Jansson, G. (2009): Haptic Interaction. In: Stephanidis, Constantine (ed.). "The Universal Access Handbook". Boca Ration, USA: Taylor and Francispp. 33-1-33-16*

*Rau, Pei-Luen and Marcus, Aaron (2009): International and Intercultural User Interfaces. In: Stephanidis, Constantine (ed.). "The Universal Access Handbook". Boca Ration, USA: Taylor and Francispp. 9-1-9-12*

*Rideout, Victoria J., Vandewater, Elizabeth A. and Wartella, Ellen A. (2005): Electronic Media Use in the Lives of Infants, Toddlers, and Preschoolers. In American Behavioral Scientist, 48 (5) pp. 501-504*

*Roberts, Vera Louise and Fels, Deborah I. (2006): Methods for inclusion: Employing think aloud protocols in software usability studies with individuals who are deaf. In International Journal of Human-Computer Studies, 64 (6) pp. 489-501*

*Savidis, Anthony and Stephanidis, Constantine (2009): Unified Design for User Interface Adaptation. In: Stephanidis, Constantine (ed.). "The Universal Access Handbook". Boca Ration, USA: Taylor and Francispp. 16-1-16-17*

*Savidis, Anthony and Stephanidis, Constantine (2004): Unified user interface design: designing universally accessible interactions. In Interacting with Computers, 16 (2) pp. 243-270*

*Sears, Andrew, Lin, M., Jacko, Julie A. and Xiao, Y. (2003): When Computers Fade... Pervasive Computing and Situationally-Induced Impairments and Disabilities. In: Stephanidis, Constantine (ed.) Proceedings of the Tenth International Conference on Human-Computer Interaction June 22-27, 2003, Crete, Greece. pp. 1298-1302*

*Shinohara, Kristen (2006): Designing assistive technology for blind users. In: Eighth Annual ACM Conference on Assistive Technologies 2006. pp. 293-294*

*Smith-Jackson, Tonya, Nussbaum, Maury A. and Mooney, Aaron M. (2003): Accessible cell phone design: development and application of a needs analysis framework. In Disability & Rehabilitation, 25 (10) pp. 549-560*

*Starner, Thad, Auxier, Jake, Ashbrook, Daniel and Gandy, Maribeth (2000): The gesture pendant: a self-illuminating, wearable, infrared computer vision system for home automation control and medical monitoring. In: Proceedings of the 4th IEEE International Symposium on Wearable Computers October 16-17, 2000, Atlanta, USA. pp. 87-94*

*Stephanidis, Constantine (2001): The concept of Unified User Interfaces. In: Stephanidis, Constantine (ed.). "User Interfaces for All - Concepts, Methods, and Tools". Mahwah, USA: Lawrence Erlbaum Associatespp. 371-388*

*Stephanidis, Constantine and Emiliani, Pier L. (1999): Connecting to the Information Society: a European Perspective. In Technology and Disability Journal, 10 (1) pp. 21-44*

*Stephanidis, Constantine and Emiliani, Pier L. (2005): Universal Access to Ambient Intelligence Environments: Opportunities and Challenges for People with Disabilities. In IBM Systems Journal, 44 (3) pp. 605-619*

*Stephanidis, Constantine and Salvendy, Gavriel (1999): Toward an Information Society for All: HCI Challenges and R&D Recommendations. In International Journal of Human-Computer Interaction, 11 (1) pp. 1-28*

*Stephanidis, Constantine and Salvendy, Gavriel (1998): Toward an Information Society for All: An International Research and Development Agenda. In International Journal of Human-Computer Interaction, 10 (2) pp. 107-134*

*Stephanidis, Constantine, Paramythis, Alexandros, Sfyrakis, Michael and Savidis, Anthony (2001): A Case Study in Unified User Interface Development: The AVANTI Web Browser. In: Stephanidis, Constantine (ed.). "User Interfaces for All - Concepts, Methods, and Tools". Mahwah, USA: Lawrence Erlbaum Associatespp. 525-568*

*Stephanidis, Constantine, Antona, Margherita, Savidis, Anthony, Partarakis, Nikolaos, Doulgeraki, Constantina and Leonidis, Asterios (2012): Design for All: Computer Assisted Design of User Interface Adaptation. In: Salvendy, Gavriel (ed.). "Handbook of Human Factors and Ergonomics (4th Edition)". John Wiley and Sonspp. 1484-1508*

*Truong, Khai N., Hayes, Gillian R. and Abowd, Gregory D. (2006): Storyboarding: an empirical determination of best practices and effective guidelines. In: Proceedings of DIS06: Designing Interactive Systems: Processes, Practices, Methods, & Techniques 2006. pp. 12-21*

*University, North Carolina State (2008). The Center for Universal Design. Retrieved 25 January 2014 from North Carolina State University: http://www.ncsu.edu/ncsu/design/cud/about_ud/udprinciples.htm*

*Vanderheiden, Gregg C., Chisholm, Wendy A. and Ewers, Neal (1996). Design of HTML pages to increase their accessibility to users with disabilities, Strategies for today and tomorrow. University of Wisconsin - Madison*

*Vanderheiden, Gregg C., Treviranus, Jutta, Gemou, Maria, Bekiaris, Evangelos, Markus, Kasper, Clark, Colin and Basman, Antranig (2013): The Evolving Global Public Inclusive Infrastructure (GPII). In: Stephanidis, Constantine and Antona, Margherita (eds.). "Universal Access in Human-Computer Interaction. Design Methods, Tools, and Interaction Techniques* *for eInclusion". Berlin, Germany: Springerpp. 107-116*

*Vredenburg, Karel, Isensee, Scott and Righi, Carol (2001): User-Centered Design: An Integrated Approach (Software quality institute series). Prentice Hall*

*W3C. Mobile Web Best Practices 1.0. Mobile Web Best Practices 1.0. Working Draft 17 October 2005.*

*W3C (1999). Web Content Accessibility Guidelines 1.0. Retrieved 25 January 2014 from W3C: http://www.w3.org/TR/WCAG10/*

*W3C (2008). Web Content Accessibility Guidelines 2.0. Retrieved 25 January 2014 from W3C: http://www.w3.org/TR/WCAG20/*

*World Health Organization (2001). International Classification of Functioning, Disability and Health (ICF). Retrieved 25 January 2014 from World Health Organization: http://www.who.int/classifications/icf/en/*

*Zabulis, Xenophon, Argyros, Antonis A. and Baltzakis, Haris (2009): Vision-based Hand Gesture Recognition for Human-Computer Interaction. In: Stephanidis, Constantine (ed.). "The Universal Access Handbook". Boca Ration, USA: Taylor and Francispp. 34-1-34-30*

*Zimmermann, Gottfried and Vanderheiden, Gregg C. (2008): Accessible design and testing in the application development process: considerations for an integrated approach. In Universal Access in the Information Society, 7 (1) pp. 117-128*

## About the author

### Constantine Stephanidis

Has also published under the name of:
"C. Stephanidis" and "Constantinos Stephanidis"

Personal Homepage:
ics.forth.gr/hci/index_main.php?l=e&c=72

Constantine Stephanidis, Professor at the Department of Computer Science of the University of Crete, is the Director of the Institute of Computer Science, Foundation for Research and Technology - Hellas, Head of the Human - Computer Interaction Laboratory, and of the Centre for Universal Access and Assistive Technologies, and Head of the Ambient Intelligence Programme of ICS-FORTH. Over the past 25 years, Prof. Stephanidis has been engaged as the Scientific Responsible in more than 40 National and European Commission funded projects in the fields of Human-Computer Interaction, Universal Access and Assistive Technologies. In the beginning of the '90s he introduced the concept and principles of Design for

All in Human-Machine Interaction and for Universal Access in the evolving Information Society. He has published more than 550 technical papers in scientific archival journals, proceedings of international conferences and workshops related to his fields of expertise. He is the Editor-in-Chief of the Springer international journal "Universal Access in the Information Society". He is the Editor and (co-) author of 14 out of the 30 chapters of the book " User Interfaces for All - concepts, methods and tools" published by Lawrence Erlbaum Associates (2001). He is also Editor and (co-)author of many chapters of "The Universal Access Handbook" that was published in June 2009 by CRC Press Taylor & Francis Group. During the period 1995 - 2006, he was the Founding Chair of the ERCIM Working Group "User Interfaces for All" and General Chair of its 9 international Workshops. During the period 1997 - 2000, he was the Founding Chair of the International Scientific Forum "Towards an Information Society for All", in the context of which he edited White Papers concerning the roadmap and R&D agenda towards an Information Society for All. Since 2001 he is the Founding Chair of the International Conference "Universal Access in Human - Computer Interaction". Since 2007 he is the General Chair of the HCI International Conference, that takes place every two years with around 2,000 participants.

## Your notes and thoughts on chapter 42

**NOTES:**

# Chapter 44

# Affordances and Design

by Victor Kaptelinin.

## 44 ABSTRACT

The concept of affordances originates from ecological psychology; it was proposed by James Gibson (1977, 1979) to denote action possibilities provided to the actor by the environment. In the late 1980s Norman (1988) suggested that affordances be taken advantage of in design. The suggestion strongly resonated with designers' concern about making possible uses of their products immediately obvious, and soon the concept came to play a central role in interaction design and Human-Computer Interaction (HCI). This chapter discusses the origins, history, and current interpretations of affordances in HCI research, and reflects on the future of affordances as an HCI concept.

## 44.1 INTRODUCTION: WHY AFFORDANCES?

Good designs are intuitive.[1] Take for instance the Holmes stereoscope, designed in the 19th century (Figure 44.1). You can immediately see that: (a) there is a handle, which you can grasp with either right hand or left hand, (b) you hold the device, so that it is supported from below, (c) you can insert stereo cards (or "stereoviews") in a card holder slot, and (d) you can view the cards through a pair of lenses. The shape of the hood surrounding the lenses indicates how exactly the device should be placed for proper viewing.

Even if you haven't seen a Holmes stereoscope before, you are likely to be able to use it almost immediately.



FIGURE 44.1: A Holmes stereoscope.

1. This chapter appeals to a commonsense understanding of "intuitive" and does not intend to discuss the exact meaning of the term. Such discussions can be found elsewhere (e.g., Raskin, 1994; Baerentsen, 2000; O'Brien et al., 2010).

There are myriads of cleverly, intuitively designed things around us, both old and new. Some examples include a car door handle, which we use correctly without thinking, even if we are encountering that particular handle for the first time (Figure 44.2), a Swiss Army knife (Figure 44.3), a summer cottage window lock (Figure 44.4), and so on. The list of things that dutifully and unobtrusively serve us in our daily lives is endless.



Figure 44.2: Intuitive everyday designs: Car door handles.



Figure 44.3: An intuitive everyday design: Swiss army knife.

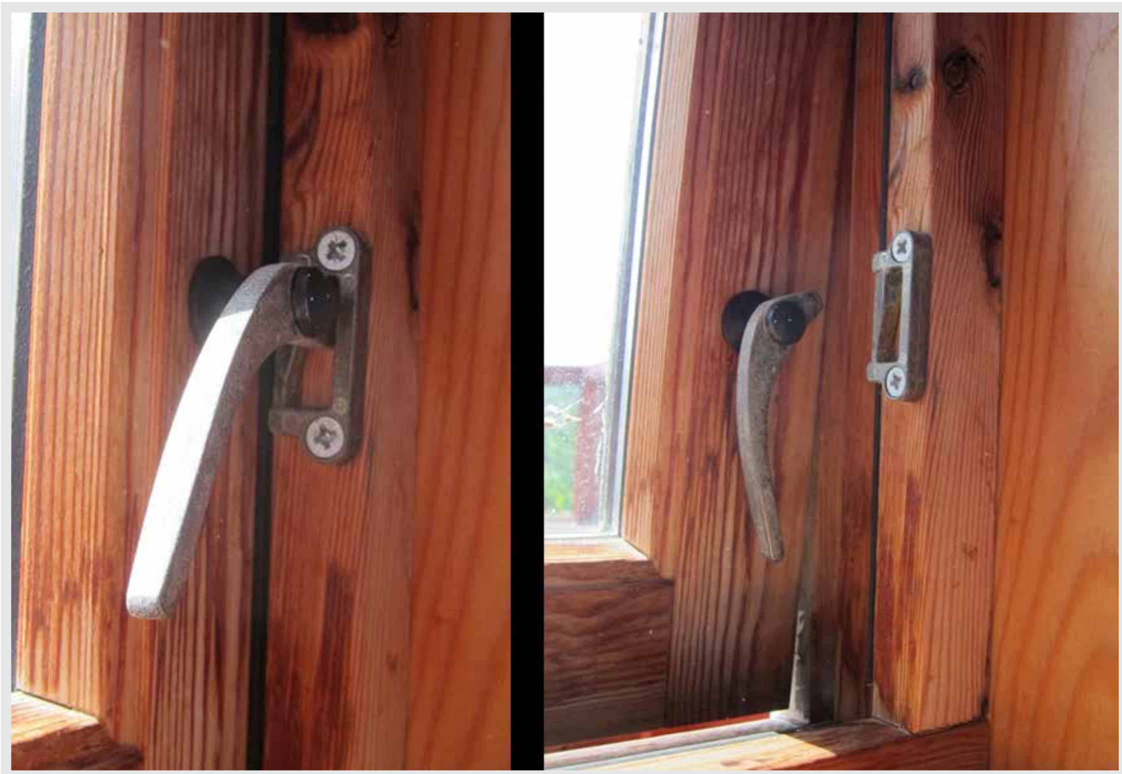




Figure 44.4: An intuitive everyday design: Window lock.

However, the design of some of the things we encounter in our daily lives is not exactly intuitive – and, unfortunately, poorly designed things are not that uncommon. An insightful discussion of a diversity of confusing and frustrating objects, such as doors that may easily turn into traps, can be found in Norman (1988).

Poor designs can even have far-reaching political consequences. Tognazzini (2001) argues that the design of the butterfly ballot used in Palm Beach, Florida, during the 2000 US presidential election, may have tipped the balance of the election as a whole. Arguably, thousands of voters were confused by the design of the ballot and voted for the wrong candidate (see Figure 44.5).

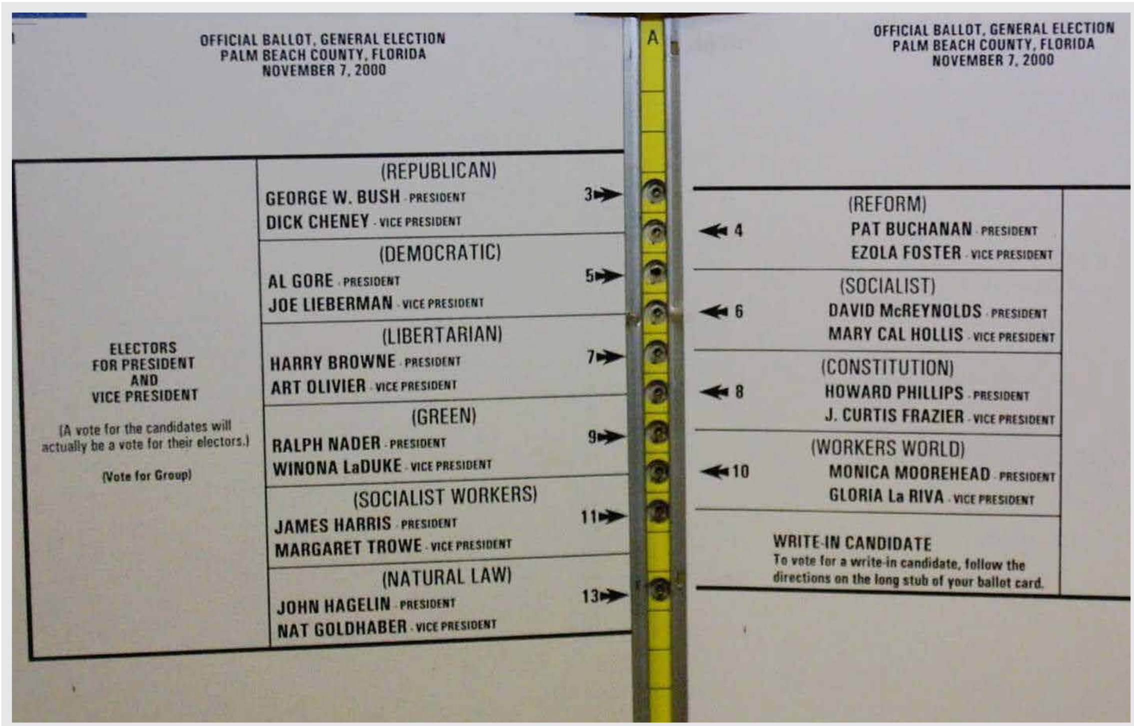





Figure 44.5: A counterintuitive design: The “Butterfly ballot”. The Democratic Party is listed second on the left column, but in order to vote for it one should press the third button. Pressing the second button would cast a vote for the Reform Party.

What is the secret of making designs intuitive? As suggested by the examples above, an essential part of it has to do with *perception*. It is not sufficient for a good design to be rational and logical. Great, intuitive designs are those that allow us directly, and correctly, *to see* what we can do with a thing.

Direct perception of possibilities for action is, essentially, what the concept of affordance is about. The concept was originally proposed by an American psychologist, James Gibson, to denote what the environment “offers the animal, what it provides or furnishes, either for good or ill.” (Gibson, 1979). The concept was introduced to the field of design, and eventually HCI, by Donald Norman in his groundbreaking book The Psychology of Everyday Things (1988). Norman defined affordances as:

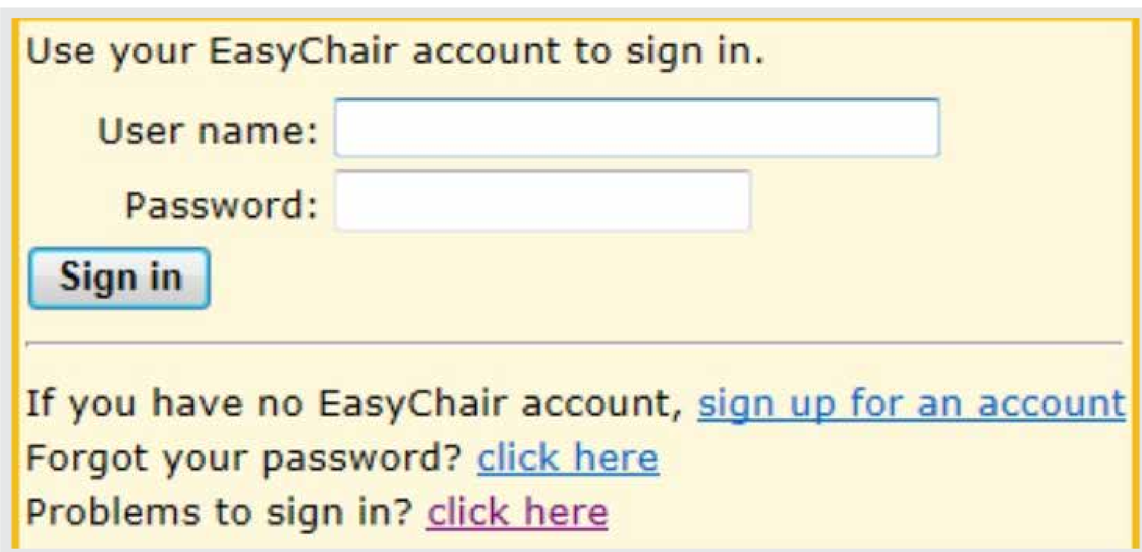
> *"... the perceived or actual properties of the thing, primarily those fundamental properties that determine just how the thing could possibly be used... A chair affords ('is for') support and therefore affords sitting. A chair can also be carried. Glass is for seeing through, and for breaking."*
>
> *(Norman, 1988).*

Affordances, according to Norman, can be fruitfully employed in design:

> *"Affordances provide strong clues to the operations of things. Plates are for pushing. Knobs are for turning. Slots are for inserting things into. Balls are for throwing or bouncing. When affordances are taken advantage of, the user knows what to do just by looking: no picture, label, or instruction needed."*
>
> *(Norman, 1988).*

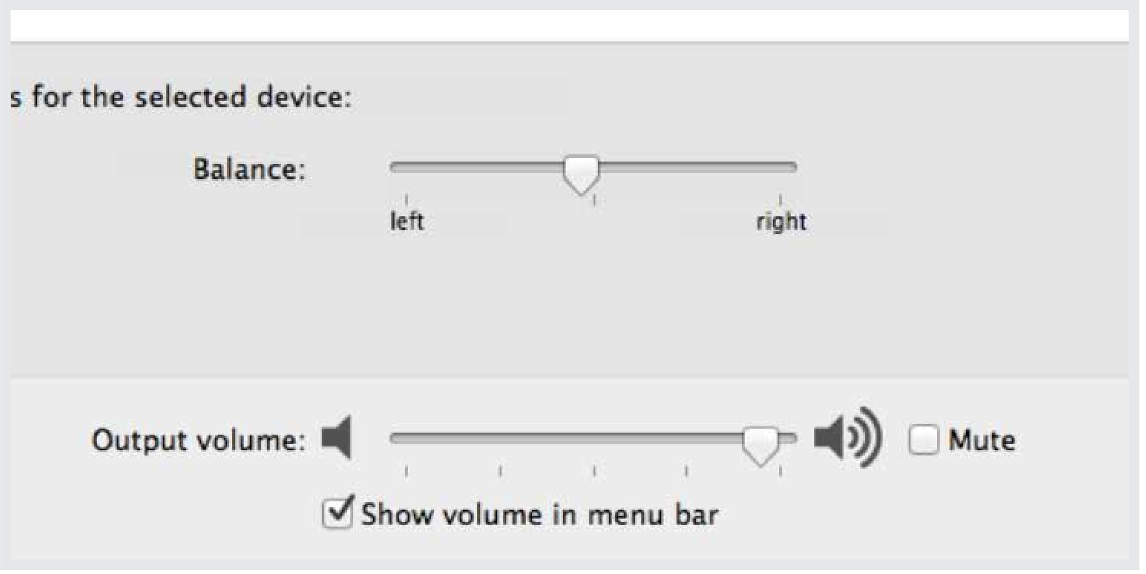
The concept of affordances was quickly adopted in HCI and interaction design; it became popular among practitioners, researchers, and educators. For designers of interactive technologies the concept signified the promise of exploiting the power of perception in order to make everyday things more intuitive and, in general, more usable. Affordance is also considered a fundamental concept in HCI research and described as a basic design principle in HCI and interaction design textbooks (e.g., Rogers et al., 2011).

The use of affordance is not limited to the design of physical objects. In fact, the concept has been especially appealing to designers of graphical user interfaces. Compared to traditional industrial designers, user interface designers can more freely and easily define visual properties of the objects they create. Therefore, they appear to be particularly well positioned for providing what Norman (1988) calls "strong visual clues to the operation of things". Examples of user interface elements, which provide this kind of strong clues, are clickable[2] buttons and tabs, draggable sliders, and spinnable controls, as well as other elements that more or less directly suggest suitable user actions (see Figure 44.6).

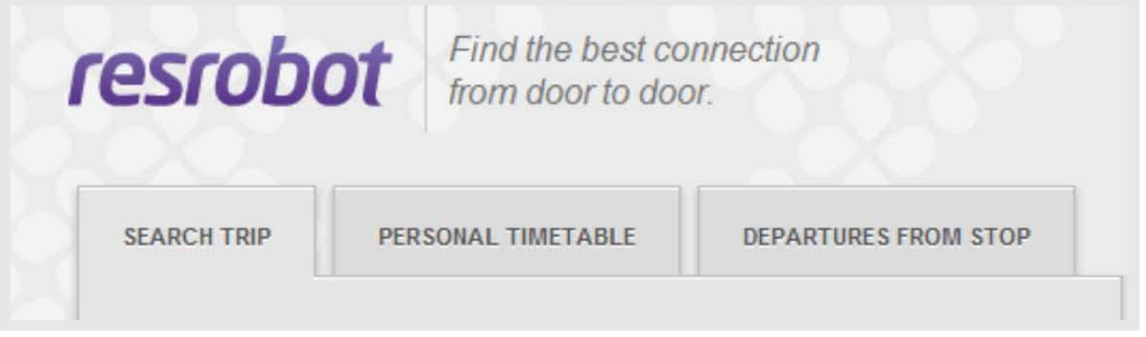





2. The notion of ambient light is contrasted to static flat pictures, which are often used in traditional psychological studies of perception.


s for the selected device:
Balance:
left
right
Output volume:
Mute
Show volume in menu bar




resrobot
Find the best connection
from door to door.
SEARCH TRIP
PERSONAL TIMETABLE
DEPARTURES FROM STOP


*Courtesy of Flickr User Jodiepedia. Copyright: CC-Att-SA (Creative Commons Attribution-ShareAlike 3.0 Unported).*




Figure 44.6: Examples of user interface elements, which directly suggest suitable user actions: (a) clickable button and hyperlinks, (b) draggable sliders, (c) clickable tabs, (d) a swipable touchscreen slider, (e) “pressable” buttons and “spinnable” controls of a touchscreen widget.

Affordance is not only one of the most central of HCI concepts, but also one of the most controversial: its history in HCI is abundant with twists and turns. The meaning of the concept and its relevance to HCI and interaction design have been subjects of debate for over two decades.

This chapter discusses the concept of affordances and how it has been used in HCI; it explores the continuing debate in HCI research concerning theoretical interpretations and design implications of the concept. The remainder of the chapter is organized into four parts:

- Theoretical roots: A brief walkthrough of the history and main points of Gibson's theory of affordances.
- Affordances in HCI research: An overview of selected analyses.
- Key issues of debate: A discussion of some of the most controversial issues.
- Conclusion: Reflections on the present and future of the concept of affordances in HCI and interaction design.

## 44.2 THEORETICAL ROOTS

This section presents a brief overview of the theoretical roots of the concept of affordances. It discusses some relevant work in ecological psychology – the field in which the concept was originally developed before it was "imported" to HCI. The main focus is on the notion of affordances, proposed by Gibson (1977, 1979), while more recent, post-Gibsonian developments in ecological psychology are only mentioned in passing. The discussion in the section is not specifically related to HCI and interaction design; its aim is to clarify the original meaning of affordances and thus provide necessary grounding for analysis in the sections that follow.

### 44.2.1 Gibson's ecological approach to visual perception

The concept of affordance was proposed by James Gibson (1977, 1979) as part of his ecological approach to visual perception. In traditional cognitive psychology perception is commonly understood as a process of developing representations. In this process sensory data that initially have no meaning, are combined with information stored in memory, interpreted, and eventually become meaningful. Gibson strongly opposed this view. He proposed an alternative, *anti-representationalist,* theory of perception.

#### 44.2.1.1 Mutuality of animal and environment

A key idea underlying Gibson's approach is *mutuality (or complementarity) of animal and environment*. Animal and environment are two parts of a whole system: one of them implies the other. There is a coupling between animals' anatomy and behavior, on the one hand, and the structure of their environments, on the other, which makes it possible for the animals to survive and successfully act in the environments. At the same time, the notion of "environment" includes, if implicitly, the animal. We do not describe our environments in terms of atoms or galaxies. Instead, we point to objects (rooms, furniture, trees, paths, streets, hills, etc.) that commensurate with us as animals of a certain size and having certain action capabilities.

#### 44.2.1.2 Detecting invariants in ambient light

The notion of mutuality of animals and environments implies that there is no particular need for animals to create a representation of the "objective world". The purpose of perception is to efficiently obtain *meaningful* information, that is, information that has significance to acting in the environment.

Four arguments are critical to Gibson's reasoning. First, he observes that environments are *structured*: they are organized into dynamically changing configurations of substances and surfaces, which comprise objects, layouts, and events. Second, these structures of the environment are *meaningful* to the animal. For instance, they can mean shelters, tools, paths, obstacles, collisions, and so forth. Third, Gibson asserts that these structures, in turn, give *structure to ambient light*, that is, light that is reflected from objects in the environment and comes to the animal from all directions. Structured ambient light, or ambient optic array, can also change from moment to moment, for instance, because the animal is changing its position.[3] Fourth, Gibson argues that by detecting invariants in ambient light, corresponding to significant aspects of the environment, animals directly pick up meaningful information without developing internal representations of their environments.

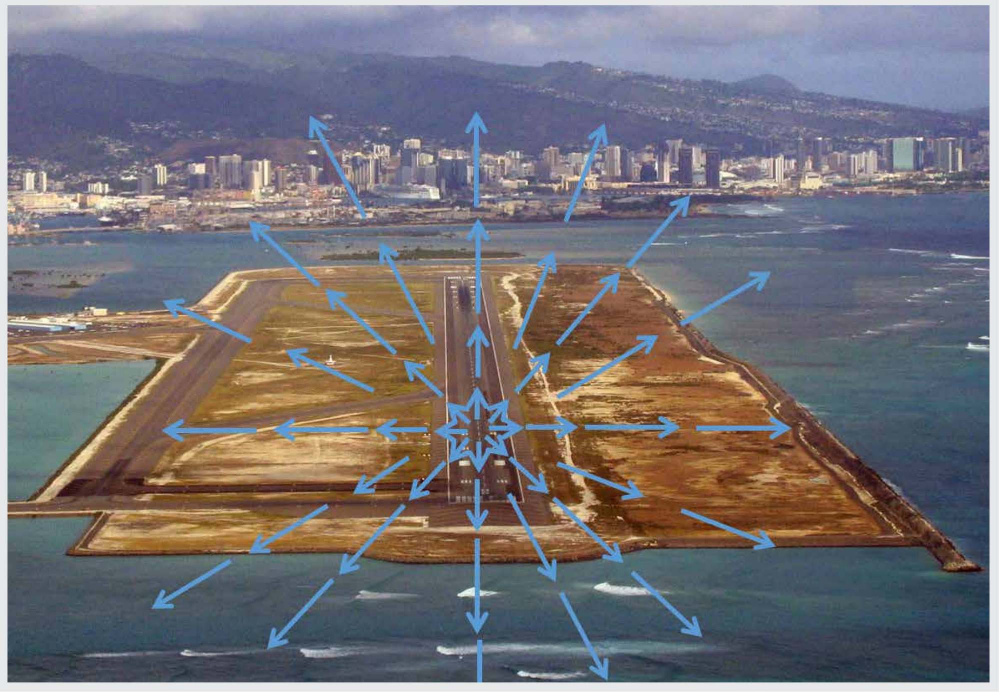




FIGURE 44.7: An outflow of the ambient optic array indicating that airplane is on the landing glide.

3. An alternative interpretation of affordances as emergent properties of the animal-environment system, rather than properties of the environment *per se*, is proposed by Stoffregen (2003). A critical discussion of this position is presented, for instance, by Kirlik (2004).

Invariants in energy arrays can be rather complex and include sensory data widely distributed in space and time. For instance, a certain pattern of how an entire ambient optic array changes over time indicates to the pilot that the plane is landing (Figure 44.7) rather than taking off or flying over the terrain.

Animal's activity in the environment, including both body part movements and locomotion, is a crucial part of detecting invariants. Birds move their heads to perceive depth: differences in visual field caused by moving the point of view help them compensate for the lack of binocular vision.

#### 44.2.1.3 The notion of affordance

What kind of meaningful information about environment do animals directly pick up from ambient light? According to Gibson, this is information about *affordances*, that is, *action possibilities offered by the environment to the animal*. Affordances are determined by both the environment and the animal (or, more specifically, action capabilities of the animal). For instance, a chair affords sitting to animals having certain bodies – in other words, for such animals it is *seatable*. A hill can be *climbable* for some animals (and *un-climbable* for others), a needle is *pierce-with-able (*for people and not, for instance, dogs*)*, and so on.

An affordance is a property of the environment; it can be measured and studied objectively. At the same time, it is a *relational* property – it is determined by the relationship between animal and environment rather than by the environment alone.[4] For the sake of illustration, let us consider a simple example (a similar example is used by Vyas et al., 2006). Imagine that it has been empirically established that sheep cannot jump over a fence if the fence's height exceeds a certain value (say, 117 centimeters, an arbitrary figure). In other words, fence's height, an objective attribute of the fence, can be used to determine whether or not the fence

4. For instance, a search in the ACM Digital Library using "HCI" & "affordances", performed on June 16, 2013, produced 1,790 hits.

in question is jump-over-able by comparing it to an empirically established, objective value (in our case, 117 centimeters). But even though it is the fence that may or may not offer the affordance, fence's affordance only exists in relation to one particular animal species, sheep. It cannot be assumed to be the same for, say, horses.

It should be specifically emphasized that Gibson was not interested in affordances per se. To him affordances were relevant only to the extent to which they could help provide an account of how animals perceive their environments. He pointed out:

> *"The central question for the theory of affordances is not whether they exist and are real but whether information is available in ambient light for perceiving them."*
>
> *(Gibson, 1979).*

Gibson's theory asserts that animals directly pick up information about affordances, which makes detection of critically important aspects of the environment quick and efficient. For instance, when we see the brink of a cliff right in front of us we directly recognize that it affords falling off and immediate action (or inaction) is needed to avoid the danger. Such direct perception appears to be quite successful. Preikestolen ("Preacher's pulpit"), a 600 meter-high cliff in Norway, is a major tourist attraction, visited each year by over 100 000 people (Figure 44.7). Even though there is no safety railing on the top of the cliff and many visitors enjoy standing or sitting close to the edge, no accidental falls have been reported so far.



Figure 44.8: Preikestolen in Norway.

The concept of affordance bears some similarity to earlier concepts proposed in Gestalt psychology. In particularly, Gibson acknowledged that his work was influenced by Koffka's notion of "demand character" and Lewin's notion of "invitation character", or "valence" (Gibson, 1979). At the same time, Gibson insisted that there was a substantial difference between these concepts and "affordance":

> *"The concept of affordance is derived from these concepts of valence, invitation, and demand, but with a crucial difference. The affordance of something does not change as the need of the observer changes. The observer may or may not perceive or attend to the affordance, according to his needs, but the affordance, being invariant, is always there to be perceived. An affordance is not bestowed upon an object by a need of an observer and his act of perceiving it. The object offers what it does because it is what it is."*
>
> *(Gibson, 1979, original italics)*

### 44.2.1.4 Cultural and natural environments

Gibson's approach does not make a fundamental distinction between human beings and other animals. The assumption of the mutuality of animal and environment, as well as the arguments based on this assumption, which lay out the foundation of the theory of affordances, are general enough to be applicable to any animal. Some examples of affordances described by Gibson are related to specifically human objects, such as mailboxes, and he paid special attention to a variety of tools, including scissors, knives, and clubs. However, these affordances are considered similar to affordances provided by "natural" objects to non-human animals. Gibson observed:

> *"...it would be a mistake ... to separate the cultural environment from the natural environment, as if there were a world of mental products distinct from the world of material products. There is only one world, however diverse, and all animals live in it, although we human animals have altered it to suit ourselves."*
>
> *(Gibson, 1979)*

### 44.2.2 Selected analyses of affordances in post-Gibsonian ecological psychology

Gibson's concept of affordance was further explored and elaborated in a number of more recent studies in ecological psychology (e.g., Heft, 2000). These studies have made a relatively limited impact on HCI research compared to the original work by Gibson, while some of them are potentially relevant to HCI. This section briefly discusses a few selected examples of such studies.

As mentioned, the discussion is not intended to be a comprehensive analysis of post-Gibsonian developments in ecological psychology from the point of view of their potential implications for HCI. Such analysis is a separate (and much needed) task, which is beyond the scope of this chapter.

#### 44.2.2.1 Learning

The notion of *learning* does not play a significant role in Gibson's original theory of affordances. Learning is briefly mentioned on a few occasions, for instance, when observing that perception can be enriched and refined with practice, but the issue of how exactly people learn to perceive a new affordance is basically avoided. Gibson acknowledges that perception of affordances can be incorrect (which is what he calls "misperception") and therefore there may be a need to un-learn existing invariants and/or learn new ones, but he maintains that basic affordances do not require much learning.

A somewhat different view on the role of learning in the perception of affordances was presented by Eleanor Gibson (James Gibson's wife and a prominent psychologist herself) and Anne Prick (2003). In their book, *An ecological approach to perceptual learning and development,* they assert that affordances do not automatically present themselves to the actor. Instead, they typically must be *discovered* through perceptual learning, and actors must *learn to use* the affordances, which in some cases "...may require much exploration, patience, and time" (Gibson and Pick, 2003, p. 17).

The studies reported by Eleanor Gibson and Anne Prick show that much perceptual learning takes place during infancy. Growth provides infants with more advanced action and sensory systems, and these new capabilities are employed by the infants to expand and differentiate their perceptual worlds. Gibson and Prick also conclude that perceptual learning and development in infants is species-typical: it is generally similar for all infants of the same species. Arguably, a limitation of the analysis presented by Gibson and Prick is that it is predominantly concerned with perception in infants and young children. While it is stated that perceptual learning continues after infancy and becomes more diverse and specific, little research on this issue is reported.

#### 44.2.2.2 Tools

To James Gibson, tools are one of the main types of meaningful objects in the environment. He mentions several kinds of tools and their affordances.

For instance, scissors are described as an extension of the human hand (Gibson, 1979). However, Gibson does not present a systematic conceptual analysis of what makes tools different from other objects in the environment. Some of the recent studies in ecological psychology provide more specific evidence on this issue.

For instance, Wagman and Carello (2001, 2003) conducted a series of studies of how people use a particular tool, a rod. Rods and sticks can be used, among other things, for hammering and poking, so their affordances include '*hammer-with-ability*' and '*poke-with-ability*'. In the experiments conducted by Wagman and Carello it was found that when a stick is intended to be used for different purposes (hammering vs. poking) different grips were employed by the participants and the use of a tool depended on how people explored inertial constraints – even without being able to see a tool. When physical parameters of a stick, e.g., those relevant to its *hammer-with-ability* (such as the relative weight of different parts of a stick) were modified, corresponding changes in the grip were observed. Wagman and Carello conclude that when analyzing how people use affordances of a tool one should differentiate between tool-user interface and tool-environment interface. They also emphasize the importance of studying how visual information is combined with perceptual information from other modalities.

#### 44.2.2.3 Collaborative action

Gibson's framework is almost exclusively concerned with how *individual* animals perceive and act in their environments (which environments may include other animals, too). But animals, especially human beings, can also perform joint, collective actions. For example, several people can carry an object, such as a stretcher, which can be too heavy or bulky to be carried by a single person. In recent research Gibson's theory of affordances was extended to such actions as well, and used in studies of how people perceive action possibilities for joint actions. For instance, Davies et al. (2010), who analyzed how people view a possibility to go through a doorway together with another person, have shown that possibilities for joint actions, – that is, actions performed by two persons at the same time, – can be perceived directly.

## 44.3 AFFORDANCES IN HCI RESEARCH: AN OVERVIEW

This section gives an overview of some of the key conceptual explorations of affordances in HCI research. The overview is unavoidably selective and incomplete. The sheer volume of HCI literature that uses the concept of affordances[5] makes it impossible to cover all relevant work. Some important analyses therefore may not be included in the discussion below. In addition, some insightful interpretations of the concept in areas just outside the scope of HCI (e.g., Chen et al., 2007; Ihara et al., 2009; Laarni et al., 2007; Suthers, 2006; Sahin et al., 2007; Zhang, 2008) are not discussed here.

The overview is organized around four main themes: (a) affordances in Ecological Interface Design, (b) specifying affordance as an HCI concept, (c) reframing affordances from non-Gibsonian theoretical perspectives, and (d) exploring alternative or complementary concepts.

### 44.3.1 Affordances in Ecological Interface Design

Ecological Interface Design (EID) is an approach in cognitive systems engineering, developed by Vicente and Rasmussen in the late 1980s and early 1990s (e.g., Rasmussen and Vicente, 1989; Vicente and Rasmussen, 1990). The framework is explicitly informed by ecological psychology, primarily by the work of Gibson and Brunswik (see, e.g., Rasmussen and Vicente, 1989). The concept of affordances

5. A systematic comparison of three types of mapping – symbolic, metaphorical, and nomic – can be found in Gaver (1986)

was adopted in EID at approximately the same time as it was introduced to design by Norman and, apparently, independently of Norman.

The main objective of EID is to create user interfaces for operators of complex industrial systems that would support efficient and safe work practices. The approach capitalizes upon Rasmussen's taxonomy of three levels of cognitive control: skill-based level, rule-based level, and knowledge-based level. The first two levels are concerned with perception and action, and control that takes place at these levels is faster, more effortless, and less error-prone than analytical problem solving associated with control at the knowledge-based level.

The aim of EID is to make sure that as much control as possible is performed at the lower levels (that is, skilled-based and rule-based levels). This aim is achieved by designing interfaces that make abstract invisible properties of the industrial processes visible and thus allow the operators to take advantage of the power of perception. Gibson's theory of affordances is primarily used in EID to explore design strategies for supporting the operator in direct perception of action possibilities in industrial control settings.

While EID is an influential approach with a good record of successful practical implementations, it has been relatively loosely related to other developments in HCI at large, especially in the last two decades. A likely reason is that the approach was specifically developed for highly structured complex industrial settings, which have eventually become a less central object of study in the mainstream HCI (which is now mainly interested in "loosely coupled domains", see Albrechtsen et al., 2001).

### 44.3.2 Specifying affordance as an HCI concept

When the general idea of affordances was introduced to design by Norman (1988), it was expressed, metaphorically speaking, in a few powerful brush strokes. The introduction was strong and convincing but not particularly detailed and, as subsequently acknowledged by Norman himself, a little imprecise. Some of more recent papers in HCI and interaction design make an attempt to clarify the meaning of affordances and relate the concept to specific agenda of HCI research and practice.

#### 44.3.2.1 Gaver (1991, 1992, 1996): Affordances vs. their perception, affordances for complex actions, and multimodality

An important early analysis of affordances in the context of HCI was conducted by Gaver (1991, 1992). In his paper "Technology affordances" (1991), which, as observed by McGrenere and Ho (2000), was the first CHI conference paper on affordances, Gaver provides an insightful, if rather succinct, discussion of a range of key issues that need to be elaborated upon in order to make affordance a useful and usable HCI concept.

First, Gaver systematically analyzes the relationship between affordances and perceptual information about affordances. He identifies four possible combinations of the presence or absence of *affordances*, on the one hand, and the presence or absence of *information about affordances*, on the other hand: perceptible affordances, false affordances, hidden affordances, and correct rejection (Figure 44.9). As noted by McGrenere and Ho (2000), Gaver's differentiation of affordances as such from perceptual information that specifies them (which is in line with the original Gibsonian meaning of the term) is somewhat different from Norman's (1988) interpretation, which combines affordances and their perception.

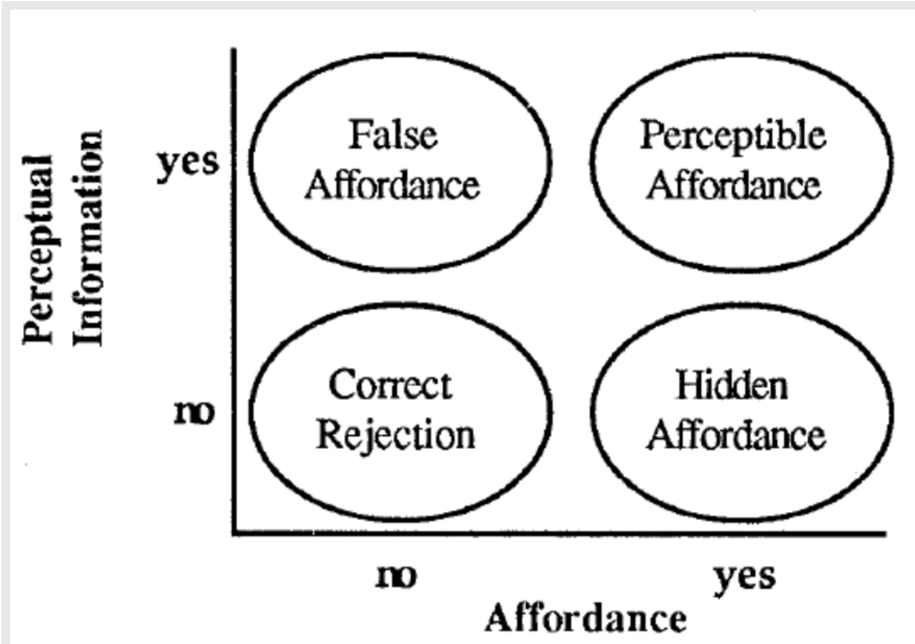





Figure 44.9: Separating affordances from the information available about them allows the distinction among correct rejections and perceived, hidden and false affordances. From Gaver (1991).

Second, Gaver discusses affordances for complex actions, that is, actions comprising several sub-actions. He identifies two types of such affordances:

- *Sequential affordances*: "acting on a perceptible affordance leads to information indicating new affordance" (Gaver, 1991, p. 82). For instance, visual information about a door handle may indicate that the handle is *graspable*, while grasping the handle may reveal that it is also *turnable*.
- *Nested affordances:* one affordance serves as context for another one. For instance, a door handle's affordance of *graspability* can be nested within the door's affordance of *pullability*.

Gaver emphasizes the importance of *active exploration* in revealing and using affordances of complex objects. He also suggests that the role of metaphors in design should be in guiding users' exploration of a system rather than conveying the actual knowledge about how exactly the system in question is supposed to be used.

Third, Gaver points out that information about affordances is not limited to visual information. Other modalities, such as tactile information and sound, as well as their combinations, are important as well, and should be taken into account in design.

In addition, Gaver briefly comments on the issue of making affordances perceptible. He observes that the attributes of the object, which are relevant for action, should be made available for perception without using mediating representations: "What is perceived is what is acted upon". Designs that successfully offer perceptible affordances, according to Gaver, are employing *nomically* (causally) mapped graphical objects, whose meaning is directly available to the perceiver.[6]

The concept of affordances has informed a number of concrete studies, conducted by Gaver, such as an investigation of how groups of people perceive and use media spaces, as opposed to regular physical spaces (Gaver, 1992; Gaver 1996).

#### 44.3.2.2 Norman (1999): Real vs. perceived affordances and types of constraints

In a paper published a decade after introducing the concept of affordances to design, Norman (1999) commented on how the concept was taken up by designers. Norman noted that employing the concept was often associated with confusion about its actual meaning, and made an attempt to clarify the confusion. In particular, he acknowledged that his interpretation was somewhat different from Gibson's original meaning, that by "affordances" he meant "perceived affordances",

6. The paper also analyzes the interpretation of affordances in Cognitive Systems Engineering, more specifically, in the work of Vicente and Rasmussen, discussed in section 3.1.1 of this chapter.

which can be different from Gibsonian “real affordances”. While the meaning of “perceived affordances” was not explicitly defined, it appears to correspond to Gaver’s (1991) “false affordances” and “perceptible affordances” (see McGrenere and Ho, 2000).

Norman uses the distinction between three types of constraints – physical, logical, and cultural – to describe the difference between “real affordances”, mental models, and conventions. He explains that real affordances are closely related to physical constraints, while good mental models go hand in hand with logical constraints, and cultural constraints are in fact conventions shared by a social group (Norman, 1999).

#### 44.3.2.3 McGrenere and Ho (2000): Degree of affordance, functional hierarchies, and usefulness vs. usability

Gaver’s work on contextualizing the concept of affordances in HCI research (Gaver 1991, 1992) was continued in a more recent paper by McGrenere and Ho (2000). The paper argues that the original Gibsonian concept of affordances needs to be further developed to become a more useful analytical tool for the design of interactive systems. Two of the directions identified by McGrenere and Ho are: (a) incorporating the notion of varying degrees of an affordance, and (b) understanding functional hierarchies of affordances.

### *Degree of affordance*

McGrenere and Ho call for moving beyond a binary view of affordance (as something that either exists or does not exist) toward a more nuanced interpretation of the “possibility for action”. In particular, it is argued that the difficulty of using an affordance is highly relevant to usability and should, therefore, be taken into account. McGrenere and Ho refer to the work of Warren (1995) as an example of research in ecological psychology that addresses this issue.

### *Functional hierarchies of affordances*

Building on Gibson's (1979) references to nested objects in the environment and Gaver's notion of nested affordances, McGrenere and Ho argue that affordances comprise functional hierarchies, not limited to physical interaction with the system:

> *"Possible actions on a computer system include physical interaction with devices such as the screen, keyboard, and mouse. But the role of affordances does not end with the physical aspect of the system [...]. The application software also provides possible actions. A word processor affords writing and editing at a high level, but it also affords clicking, scrolling, dragging and dropping. The functions that are invoke-able by the user are the affordances in software."*
>
> *(McGrenere and Ho, 2000).*

They also observe that:

> *"It is important to note that affordances exist (or are nested) in a hierarchy and that the levels of the hierarchy may or may not map to system functions."*
>
> *(McGrenere and Ho, 2000).*

In addition, McGrenere and Ho (2000) argue strongly for separating affordances from their perception (the position they ascribed to Gibson) because, as they claim, the separation would help researchers and practitioners to differentiate more clearly between two aspects of design, namely: designing the utility of an object (an affordance) and designing usability (the information that specifies the affordance). A similar position was also expressed by Tornvliet (2003). (This view on the issue of the relationship between affordance and perception is discussed in more detail in section 4.1 below).

### 44.3.2.4 Hartson (2003): Types of affordances and Norman's model of action

Norman (1986, 1988) describes the structure of human action as an execution-evaluation cycle comprising seven stages: (1) setting a goal, (2) developing an intention to act, (3) planning a sequence of actions, (4) executing the sequence of actions, (5) perceiving the state of the world caused by the execution of the action sequence, (6) interpreting the perception, and (7) evaluating the interpretation. If the goal is achieved, the action is completed. If not, the cycle is repeated over again or the action is terminated. The model makes the task of design or evaluation more manageable by breaking it down into separate components and allowing the analyst to focus on individual stages, as well as concrete relations between the stages. The model suggests that key concerns of interaction design should be bridging the gulf of execution (stages (2) – (4)) and the gulf of evaluation (stages (5)-(6)).

Hartson (2003) argues that Norman's model of action can be used to make the notion of affordances more specific and applicable in the context of design. He differentiates between four kinds of affordances: cognitive, physical, sensory, and functional. These are defined as follows:

> *"We have named the different kinds of affordances for the role they play in supporting users during interaction, reflecting user processes and the kinds of actions users make in task performance. Norman's perceived affordance becomes cognitive affordance, helping users with their cognitive actions. Norman's real affordance becomes physical affordance, helping users with their physical actions. We add a third kind of affordance that also plays an important role in interaction design and evaluation, sensory affordance, helping users with their sensory actions. A fourth kind, functional affordance, ties usage to usefulness. We offer guidelines for considering these kinds of affordance together in a design context."*
>
> *(Hartson, 2003, p.316, original italics).*

These four types of affordances are mapped to Norman's model of action: a need for cognitive and sensory affordances is located at the step of moving from an intention to act to planning a sequence of actions, physical and sensory affordances are related to the execution of the action sequence, sensory affordances are associated with perceiving the state of the world, and cognitive affordances are claimed to be needed when interpreting the perception (Figure 44.10).

A modified version of Norman's action model, called "the Interaction Cycle", is used by Hartson as a high-level organizing scheme for the User Action Framework (UAF). UAF includes structured comprehensive sets of specific usability issues, related to each kind of affordances, – with the exception of functional affordances. Functional affordances are considered a special case: they are related to a system's reactions to user actions, which reactions (or "outcomes") are often not directly visible to the user. According to Hartson, providing users with feedback on the outcomes of their actions is a special task in the design of interactive systems.

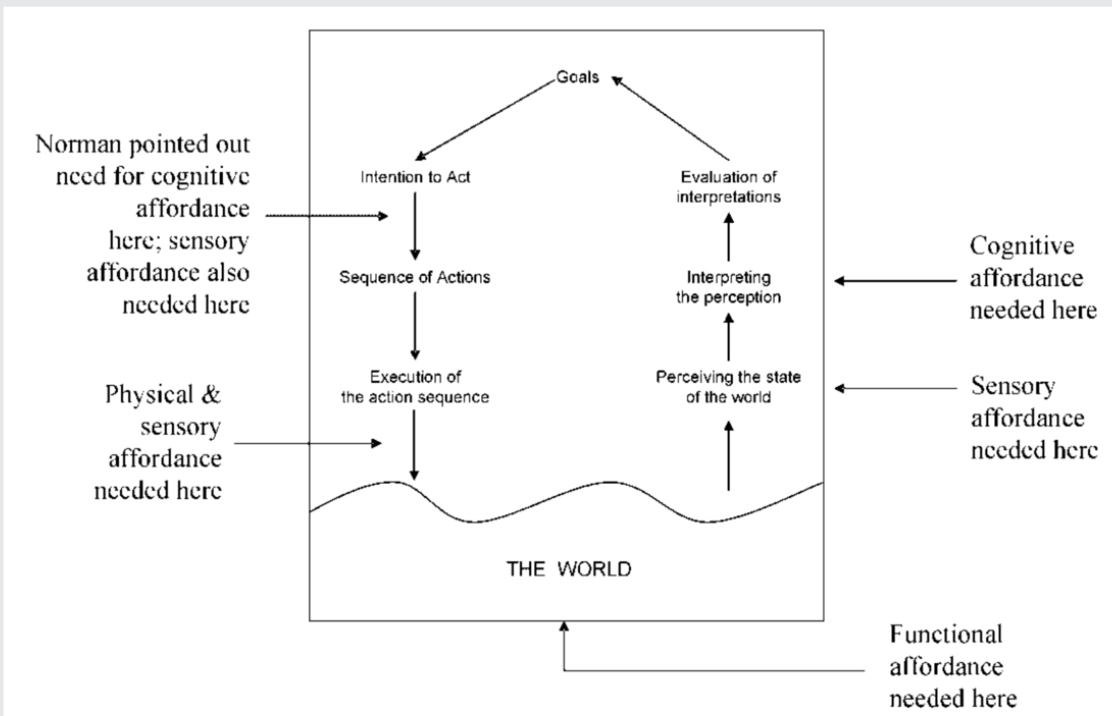





Figure 44.10: Four types of affordances mapped to Norman's action model. (Hartson 2003, p. 328).

### 44.3.3 Reframing affordances from non-Gibsonian theoretical perspectives

When limitations of information-processing psychology as a theoretical foundation of HCI became apparent to the research community (Carroll, 1991), a number of alternative approaches were adopted in the field, with activity theory and phenomenology being the leading “post-cognitivist” HCI frameworks (e.g., Boedker, 1991; Dourish, 2001; Rogers, 2004, Rogers, 2012; Kaptelinin and Nardi, 2006).

Both Heidegger’s phenomenology (Heidegger, 1962), which emphasizes the primacy of human existence in the world, and Leontiev’s activity theory (Leontiev, 1978), in which purposeful, social, mediated, and developing activity is used as a foundational concept, are similar to Gibson’s ecological psychology in postulating, in their own ways, the mutuality of the actor and the environment.

The notion of mutuality is expressed, for instance, in Heidegger’s concept of “being-in-the-world” and Leontiev’s concept of activity, which integrates “subject” and “object” in a single unit of analysis. Even though these theories do not use the term “affordance”, they all assume that perception and action are tightly integrated with one another, and the general idea of direct perception of possibilities for action fits well with their general lines of reasoning.

These approaches are also substantially different from Gibson in that they intend to move beyond animal-environment interaction and provide an account of characteristically human activities and experiences. The meaning of “possibilities for action offered by the environment” in these approaches is different from how it was understood by Gibson. Therefore, it is hardly surprising that a number of attempts have been made to reframe the concept of affordance and propose interpretations informed by activity-theory (Albrechtsen et al, 2001; Baerentsen and Trettvik, 2002; Kaptelinin and Nardi, 2012), phenomenology (Dourish, 2001; Turner, 2005; Bonderup Dohn, 2009), and some other approaches (Vyas et al., Rizzo, 2006; Rizzo et al., 2009; Still and Dark, 2013). The main points of these theoretical accounts are summarized below.

#### 44.3.3.1 Activity-theoretical accounts

##### 44.3.3.1.1 *ALBRECHTSEN ET AL. (2001): AFFORDANCES IN ACTIVITY THEORY AND GIBSON'S ECOLOGICAL PSYCHOLOGY*

A discussion of affordances from the point of view of activity theory is presented by Albrechtsen et al. (2001)[7]. The authors conclude, in particular, that there are some similarities between activity theory and Gibson's ecological psychology:

> *"Activity theory and Gibsonian thinking share the basic idea that perception is not afferent, that it is connected with action. Only through acting do people perceive their environment."*

At the same time, it is argued that activity-theory provides a broader perspective on perception and action than the Gibsonian approach. Activity theory, as opposed to Gibson's ecological psychology, is concerned with the social-historical dimension of an actor's interaction with the environment, and takes into consideration mediation and learning. Activity theory aims to provide an account of human activities at all hierarchical levels, while Gibsonian analysis generally focuses on the level of operations (using activity theory terminology). In addition, activity theory offers an understanding of tools as functional organs, a concept which does not have a counterpart in the theory of affordances. Finally, it is noted that Bødker's (1991) distinction between three complementary aspects of the use of computing technologies –physical (directed at the computer as

7. An earlier paper by Baerentsen (2000) in a special issue of the Scandinavian Journal of Information Systems on activity theory was one of the very first attempts to employ both activity theory and Gibson's theory of affordances in the context of HCI. The paper explored the notion of intuitive interfaces by using insights from both ecological psychology and activity theory. It did not, however, intend to provide a systematic account of affordances from an activity theory perspective.

a physical artifact), handling (directed at computer application), and subject/object-directed (interaction with subjects and objects through the artifact) – can be used to identify three dimensions, or types, of affordances, corresponding to the above aspects.

#### 44.3.3.1.2 *BAERENTSEN AND TRETTVIK (2002): TOWARD A CONCEPT OF AFFORDANCES BASED ON A MORE DEVELOPED NOTION OF ACTIVITY*

Another analysis of affordances from an activity-theoretical perspective was proposed by Baerentsen and Trettvik (2002).[8] Baerentsen and Trettvik start with observing that many interpretations of affordances in HCI research are deviating from the basic assumptions underlying the concept in the Gibsonian approach. They note:

> *"The concept of affordance was meant to cut though the subjective-objective dichotomy of traditional psychology and philosophy, but its interpretation in HCI often retained this dichotomy." (Baerentsen and Trettvik, 2002).*

At the same time, the authors point to some shortcomings of Gibson's theory of affordances. The main obstacle to a more successful application of the theory in HCI, according to Baerentsen and Trettvik, is an undifferentiated notion of activity employed by Gibson, which:

8. An earlier paper by Baerentsen (2000) in a special issue of the Scandinavian Journal of Information Systems on activity theory was one of the very first attempts to employ both activity theory and Gibson's theory of affordances in the context of HCI. The paper explored the notion of intuitive interfaces by using insights from both ecological psychology and activity theory. It did not, however, intend to provide a systematic account of affordances from an activity theory perspective.

> *"…makes it a difficult and nontrivial matter to address areas of research like HCI that have substantial cultural, symbolic, and technological components of a cultural-historical origin. … It is necessary to extend the analysis of affordances and their basis in organismic activity to the cultural-historical development of human activity…"*
>
> *(Baerentsen and Trettvik, 2002).*

Baerentsen and Trettvik (2002) argue that adopting a more advanced notion of activity, developed in activity theory (Leontiev, 1978), can help understand affordances as embedded in cultural contexts and emerging in concrete interaction between the actor and the environment. They identify several issues that should be taken into account in order to understand culturally-specific affordances, such as learning and the use of symbols and representations.

#### 44.3.3.1.3 *KAPTELININ AND NARDI (2012): A MEDIATED ACTION PERSPECTIVE ON AFFORDANCES*

In activity theory, tools have a special status. Human action is considered fundamentally mediated (Leontiev, 1978), and the notion of interactive technologies being mediating artifacts, through which human beings interact with the world, has informed a number of HCI concepts, models, and concrete studies (e.g., Bødker, 1991; Nardi, 1996; Beaudouin-Lafon, 2000; Bødker and Andersen, 2005; Kaptelinin and Nardi, 2006), including, as mentioned, activity-theoretical analyses of affordances (Albrechtsen et al., 2001; Baerentsen and Trettvik, 2002).

Building on this research, as well as some relevant post-Gibsonian studies in ecological psychology (e.g., Wagman and Carello, 2001; 2003), Kaptelinin and Nardi (2012) propose a mediated action perspective on affordances in HCI. They describe the structure of instrumental affordances (Figure 44.11) as comprising handling affordances (possibilities for interacting with the artifact in question) and effecter affordances (possibilities for employing the artifact to make an effect on an object of interest). For instance, a computer mouse affords moving it on a horizontal surface (handling affordance), which causes changing the pointer's position on the computer screen (effecter affordance). According to Kaptelinin and Nardi, in addition to instrumental affordances, artifacts can also provide auxiliary affordances, such as maintenance, aggregation, and learning affordances.

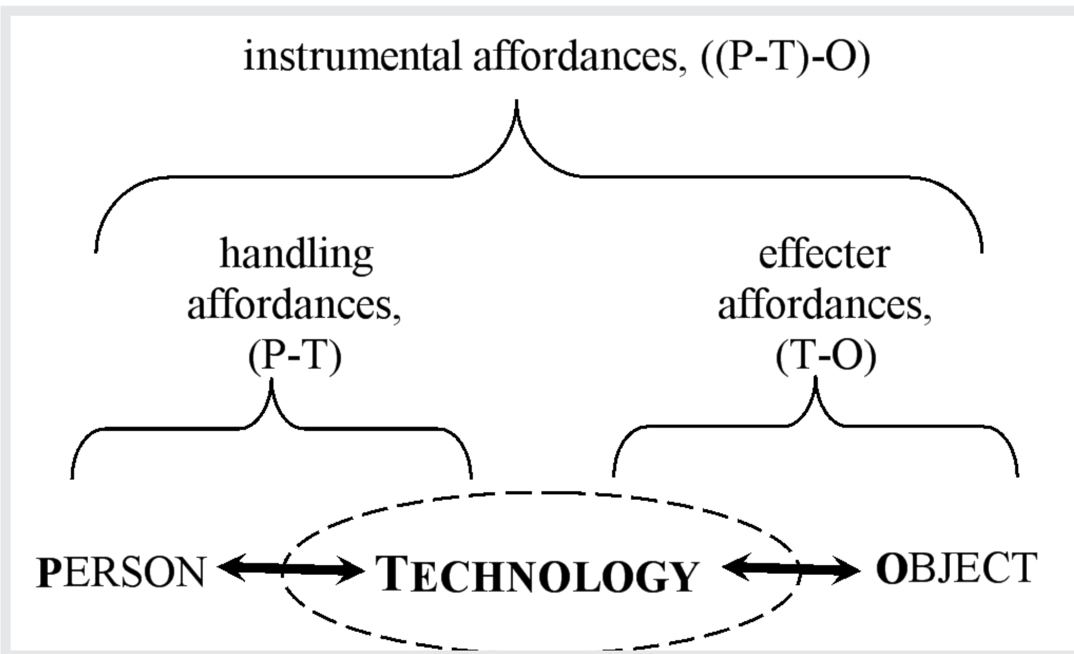





Figure 44.11: Two facets of instrumental technology affordances: handling affordances and effecter affordances- P: Person, T: Technology, O-Object of interest (Kaptelinin and Nardi, 2012).

Kaptelinin and Nardi observe that the mediated action perspective on affordances, which they are advocating, shares some basic assumptions with the original Gibsonian approach: both consider affordances as a relational property and emphasize the importance of direct perception of affordances for successful acting in the environment. At the same time, the mediated action perspective is different in a number of respects from Gibson's approach. Technology affordances are understood as relational properties emerging in a three-way interaction between actors, tools, and cultural environments. The perspective also highlights the importance of taking into account learning as well as the dynamics of a person's action capabilities, caused by tool switching. In this respect, the mediated action perspective is similar to the phenomenological account of affordances by Bonderup Dohn (2009), which is discussed in the next section.

### 44.3.3.2 Phenomenological accounts

#### 44.3.3.2.1 *DOURISH (2001): AFFORDANCES AND EMBODIED INTERACTION*

In Dourish's embodied interaction framework (Dourish, 1991), which is strongly and explicitly informed by phenomenology, the concept of affordances is used to illustrate some of the key aspects of the framework. Applications of Gibson's ecological psychology in the context of HCI and Computer Supported Cooperative Work (CSCW) are considered an example of research that is actually exploring the idea of embodied interaction. In particular, Dourish refers to how analysis and design of cooperative systems by Gaver and his colleagues (1992, 1995) incorporated the idea of an actor's exploration of the world.

"Ontology", a key aspect of meaning within Dourish's framework, is mostly discussed in relation to affordances. It is argued that the scope of the concept of affordances could be extended beyond physical actions to include affordances for particular ways of understanding the design of an artifact.

Even though the meaning of the concept of affordances is not a central issue within the embodied interaction framework[9], the use of the concept indicates that the concept is generally consistent with the phenomenological perspective.

#### 44.3.3.2.2 *TURNER (2005): SIMPLE VS. COMPLEX AFFORDANCES, SIGNIFICANCES, AND EQUIPMENT*

A deliberate attempt to conceptualize affordances from a phenomenological perspective is made by Turner (2005). Turner analyzes a variety of uses of the term "affordance" in current research and observes that the interpretations of affordances in HCI and some other fields have moved far beyond Gibson's original account.

According to Turner, current interpretations of affordances can be divided into two general categories: "simple affordances" and "complex affordances". "Simple affordances" are affordances in the Gibsonian sense of the term. "Complex affordances" are defined in terms of culture, history, and practice, and therefore cannot be properly addressed within Gibsonian ecological psychology.

Turner briefly outlines two theoretical perspectives, which he posits are capable of dealing with complex affordances. The first one is the concept of "the ideal", proposed by the Russian philosopher Evald Ilyenkov. Ilyenkov (1977) understands "the ideal" as objectively existing in the world in the form of *significances*, produced by purposeful human activities. In this respect, according to Turner, significances are similar to affordances.

The second perspective is Martin Heidegger's phenomenology (Heidegger, 1962). Turner argues that several concepts proposed by Heidegger can be used to understand complex affordances. In particular, Turner mentions Heidegger's notions of breakdowns and resulting transition of tools from being ready-to-hand to being present-at-hand, familiarity, and, especially, equipment. He also refers to a more elaborated taxonomy of breakdowns, developed by Dreyfus (2001).

9. Dourish defines affordance as a "three-way relationship between the environment, the organism, and an activity" (p. 118) but does not elaborate upon this definition.

According to Turner, since Heidegger understood equipment as context, applying Heidegger's framework to affordances leads to the conclusion that "affordances and context must be synonyms" (p. 12). This conclusion is claimed to be consistent with considering affordances as Ilyenkov's significancies.

#### 44.3.3.2.3 *BONDERUP DOHN (2009): BODY SCHEMA, DYNAMIC AND CULTURE-RELATIVE VIEW OF AFFORDANCES*

Another analysis of affordances from a phenomenological perspective is presented by Bonderup-Dohn (2009), who proposes a "dynamic, relational, and culture- and skill-dependent view" of affordances. The analysis is specifically oriented toward the field of Computer-Supported for Collaborative Learning (CSCL) but it draws heavily on the affordances debate in HCI.

Bonderup-Dohn points to the notion of "body schema"[10], proposed by the French phenomenological philosopher Merleau-Ponty (1962), as being directly relevant to understanding affordances. She observes that the notion, which emphasizes a pre-reflective correspondence of the body and the world in a concrete activity and serves as a basis for structuring the space around us and making intuitive sense of spatial relations between objects, highlights some aspects of our interaction with the world that are essential to analyzing affordances.

Of key importance to understanding technology affordances, according to Bonderup Dohn, is that body schema is a dynamic entity. Not only does it shape our interactions with the world, it is also shaped as a result of such interactions. Bonderup Dohn notes that technology can transform the body schema in a way, similar, for instance, to the one highlighted by the activity-theoretical notion of functional organs (e.g., Kaptelinin and Nardi, 2006). For instance, for a skilled typist the keyboard may become a part of the phenomenal body and "for very experienced avatar users the avatars may become incorporated into the body schema" (Bonderup Dohn, 2009).

10. The concept is also translated to English as "body image"

Adopting a Merleau-Pontian view on affordances, according to Bonderup Dohn, means that action capabilities[11] of actors should be considered depending on actors' culture and experience. Accordingly, the understanding of affordances as being culture- and experience-independent (as argued, for instance, by McGrenere and Ho, 2000) is rejected and a culture- and skill-relative interpretation of affordances is proposed instead.

### 44.3.3.3 Some other relevant theoretical accounts

Vyas et al. (2006) propose a conceptualization of affordances, according to which affordances emerge in activities and practices and are being socially and culturally constructed:

> *"...during the user-technology interaction, users actively interpret the situation and make sense of the technology while being involved in certain activities. Users' 'active interpretation' is central to the emergence of affordance that is socially and culturally determined."*
>
> *(Vyas et al., 2006).*

It is claimed that affordances should be analyzed at two levels: the artifact level and the practice level. To analyze affordances at the practice level and understand them in a broader socio-cultural context the authors suggest using Giddens' (1994) structuration theory. In addition, the framework they propose differentiates between two types of affordances: affordance in information (i.e., *what* is afforded)

11. Bonderup Dohn (2009) observes that "interaction potential" could be a more appropriate term than "action capabilities", since environments can offer the actor what she calls "intransitive affordances", that is, possibilities for actions that are not carried out by the actor, but rather by somebody else on the actor (consider, e.g., the affordance of "being seen by someone")

and affordance in articulation (i.e., *how* the system in question is supposed to be used). The ideas are further developed in a subsequent paper (Vyas et al., 2008), which differentiates between three levels of analysis of affordances: single user, organizational/work group, and societal.

Rizzo (2006) points to neurophysiological findings (Rizzolatti and Craighero, 2004) that indicate that certain basic neuronal responses correspond to whole classes of human actions having a shared goal, and similar responses can be registered when people observe other human beings trying to achieve the same goal. It is concluded that by way of imitative learning children can come to understand the action potential of objects in terms of what goals can be achieved by using these objects. Rizzo argues that it is important to understand how people communicate intentions and how "intentional affordances" (a term, originally proposed by Tomasello, 1999) are produced and perceived. Such analysis, according to Rizzo, opens up new ways to study affordances as being culturally determined through individual history. A follow up paper by Rizzo et al. (2009) suggests that to fully exploit the heuristic potential of the notion of affordances interaction design research needs to focus on the interplay between basic, sensory-motor affordances, on the one hand, and intentional affordances, on the other hand.

Still and Dark (2013) offer an account of affordances in terms of traditional cognitive psychology concepts and models. The most central concept within this account is *automatization*, the process during which the cognitive pattern recognition system learns to automatically identify constraints (irrespective of whether the constraints are physical, cultural, logical, etc.). The emergence of perceived affordances is linked to the transition from controlled to automatic processing. In cognitive psychological research automatic processing is described as "not open to awareness, rendered without intention, carrying light long-memory load, and leading to rapid responses" (p. 293). The characteristic features of controlled processing are, generally, opposite. Designers can support the transition from controlled processing to automatic processing and, therefore, help the user take advantage of perceived affordances, by making sure their designs are highly consistent.

## 44.3.4 Exploring alternative and complementary concepts

### 44.3.4.1 Norman (1999; 2008; 2011): Constraints and signifiers

As mentioned earlier, a decade after introducing the concept of affordances to HCI Donald Norman felt obliged to clarify his understanding of affordances and warn against overusing (and abusing) the concept in design (Norman, 1999). Yet another decade later he made an even more radical claim and suggested that designers should be concerned about "signifiers, not affordances". His new message to designers was: "Designers of the world: forget affordances, provide signifiers" (Norman, 2008).

The concept of signifiers, and how signifiers are different from affordances, are discussed in more detail in a more recent book entitled *Living with Complexity*, in which Norman defines signifier as "... any perceivable sign of appropriate behavior, whether intentional or non-intentional" (Norman, 2011, p. 227). He observes:

> *"It is a powerful tool for designers to enable communication. .... Signifiers act like a natural part of the world, and so the communication can be effortless and appropriate."*
>
> *(Norman, 2011, p. 228).*

Special emphasis in the introduction of the concept of signifiers is made on how signifiers are related to affordances. According to Norman, these two concepts should not be confused with one another. He notes that the concept of affordance is often misunderstood and inappropriately used by designers:

> *"(Designers…) would say that they had 'put an affordance on a product' when in fact they were making visible the presence of an already existing affordance. What they really were doing was adding a signifier."*
>
> *(p. 229).*

Norman concludes with the following claim:

> *"I strongly encourage the design community to distinguish between affordances and signifiers. In most cases, the word affordance should go away, for invariably the designer cares only about what can be perceived, which means signifiers."*
>
> *(p. 229).*

#### 44.3.4.2 DJAJADININGRAT ET AL. (2002) AND VERMEULEN ET AL. (2013): FEEDFORWARD

Vermeulen et al. (2013) discuss another concept, *feedforward*, which, as they suggest, can in certain cases be used instead of affordances. They adopt the concept of feedforward from Djajadiningrat et al. (2002), who define it as information provided to the user *before* he or she carries out the action (as opposed to *feedback*, which is information provided *after* the user carries out the action). According to Djajadiningrat et al. (2002): “Feedforward informs the user about what the result of his action will be.”

Vermeulen et al. (2013) argue that the exact meaning of feedforward is not well defined, and this is one of the reasons why the potential of this concept in design is currently underexplored. They set out to clarify the meaning by clearly separating feedforward from feedback and affordances. They capitalize upon the taxonomy of affordances and the mapping of various types of affordances to the Norman’s Stages of Action model, proposed by Hartson (2003). It is claimed that some of the elements in Hartson’s diagram (see Figure 44.9) can in fact be classified as examples of feedback and feedforward, rather than affordances. A revised diagram is suggested, which is shown in Figure 44.12.

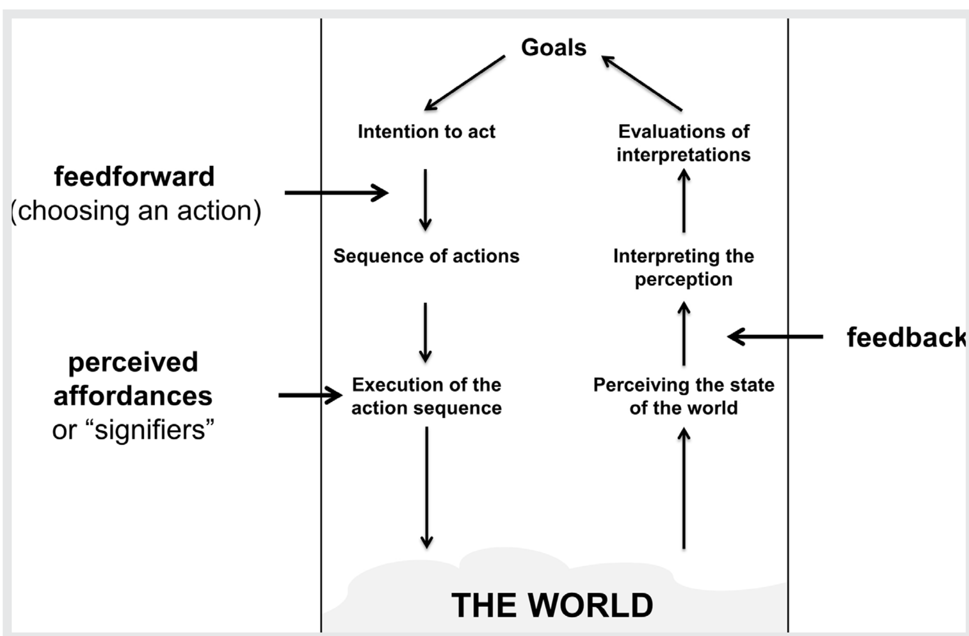





FIGURE 44.12: The position of perceived affordances, feedforward and feedback in Norman’s Stages of Action model according to Vermeulen et al. (2013).

The nature of the difference between (perceived) affordances and feedforward is described by Vermeulen et al. (2013) as follows:

> *"Both perceived affordances and feedforward tell users something about a particular action through a combination of a physical and functional affordances. Perceived affordances and feedforward essentially provide different information about the action that users have to perform to achieve their goals. While perceived affordances reveal the physical affordance, which tells users that there is an physical action available and how to perform it, feedforward reveals the functional affordance, which tells users what will happen when they perform that action."*
>
> *(Vermeulen et al., 2013, original italics).*

This distinction suggests that "user actions" as they are understood by Vermeulen et al. (2013) are, essentially, physical actions.

#### 44.3.4.3 Skeuomorphism

A design concept that is popular among interaction design practitioners (not so much among HCI researchers…) and often considered as related to affordances, is *skeuomorphism*. Generally speaking, a skeuomorph is an object or feature copying the design of a similar artifact in another material (Oxford English Dictionary, n.d.). A wallpaper pattern that copies the look of a brick wall is an example of a skeuomorph. Another example, shown in Figure 44.13, is a box for stereo cards (intended to be used with the Holmes stereoscope shown in Figure 44.1), which looks like a two-volume book set.

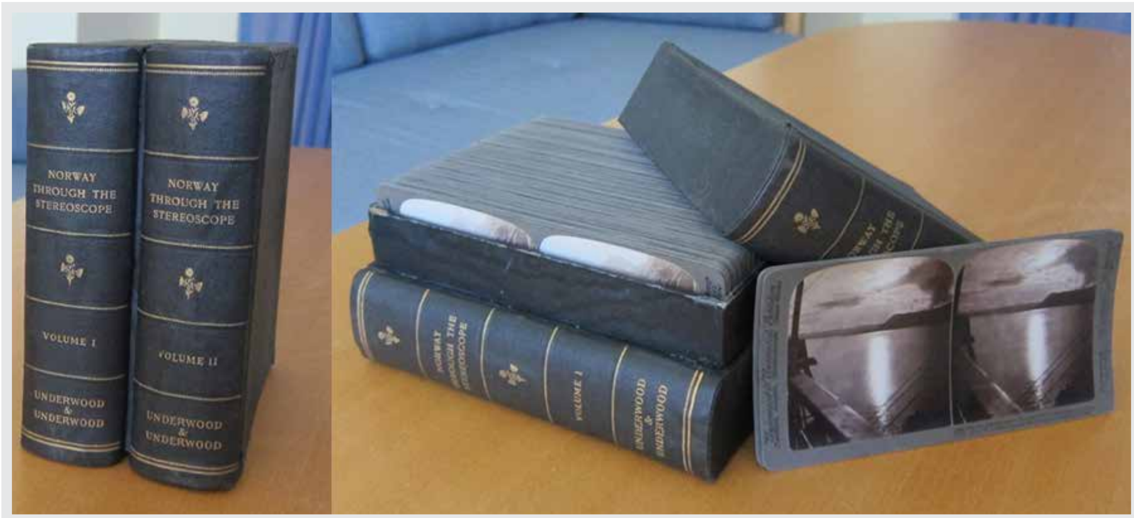





FIGURE 44.13: Skeuomorphism: A stereo card box looking like a two-volume book set).

In digital design skeuomorphism usually means realistic imitation of real-world objects, either in appearance (e.g., a stitched leather look of an electronic calendar, Figure 44.14) or in other modalities (e.g., a shutter-click sound produced by digital cameras). Skeuomorphism used to be especially common in the design of Apple products. The arguments in favor of skeuomorphism are that it makes digital objects more aesthetically pleasing and helps the user understand how to handle an unfamiliar object (which can be considered as providing perceptual information specifying object's affordances).

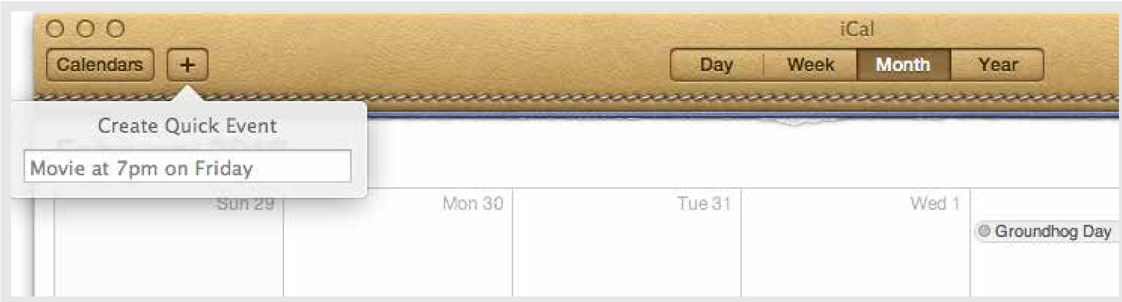





Figure 44.14: Skeuomorphism: Stitched leather look of an electronic calendar.

Recently, skeuomorphism has been losing ground in interaction design. For instance, a Technopedia.com article (Technopedia, n.d.) notes that:

> *"... skeuomorphism has increasingly come under fire, largely because many of the nostalgic elements it attempts to portray - such as calendars, day planners, address books, etc. - are almost entirely foreign to younger generations of users. In addition, critics of skeuomorphism point to this reliance of physical objects in design as an impediment to making more useful designs."*
>
> *(Technopedia, n.d.).*

The design of the latest releases of some of the most popular digital environments, such as Windows 8 and iOS 7, shows a clear trend of moving away from skeuomorphism. As a recent BBC News Magazine article observes:

> *"Skeuomorphism has fallen out of favour in recent years, and is almost regarded as a dirty word by many in the design community."*
>
> *(Judah, 2013).*

## 44.4 KEY ISSUES OF DEBATE

The brief overview of the affordances debate in HCI research, presented in the previous section, allows us to identify some common issues emerging from the debate. It should be noted that many of these issues are closely related to – and even overlap with – one another.

### 44.4.1 Affordances and perception

The relationship between affordances and perception has been a debated issue in HCI research for over two decades, with a general trend being toward progressively stricter separation of affordances from perception. This trend is especially apparent in the evolution of Norman's interpretations of affordances, discussed in detail in Section 3 above. The evolution can be briefly presented as follows[12]:

- **1988:** Norman introduces affordances to design, describing them as "perceived and actual properties of a thing"; the concept is understood as referring to both the possibilities for action, provided to the actor, and their perception by the actor;

12. It is claimed that the theory of affordances can be applied to the perception of language as well (Gibson, 1979), but the logical and empirical arguments supporting this claim are not as advanced and thorough as those provided for the direct perception of physical objects, layout, and events.

- **1999:** Norman differentiates "real affordances" (which correspond to Gibsonian "affordances") from "perceived affordances" (which may or may not be real); he clarifies that in his previous work by "affordances" he actually meant "perceived affordances";
- **2008/2011:** Norman takes a step further and completely separates affordances (which can only be affordances in the Gibsonian sense, or "real" affordances) and information about them (i.e., signifiers).

Inconsistency between Norman's initial interpretation of affordances (Norman, 1988) and the original Gibsonian meaning of the term was noticed, discussed, and found problematic by several researchers, e.g., McGrenere and Ho (2000) and Tornvliet (2004). Soegaard (2009) observes:

> *"Unlike Norman's inclusion of an object's perceived properties, or rather, the information that specifies how the object can be used, a Gibsonian affordance is independent of the actor's ability to perceive it."*
>
> *(Soegaard, 2009)*

Undoubtedly, these efforts aiming to clarify the difference between Norman's and Gibson's interpretations should get credit for resolving some terminological uncertainties. Such clarifications are important, since variations of early Norman's interpretations of affordances, abandoned by Norman himself, can still be found in literature. For instance, a popular interaction design textbook describes affordance as the term, "...which is used to refer to an attribute of an object that allows people to know how to use it." (Rogers et al., 2011).

At the same time, some attempts to clarify terminological problems go to the point of advocating the need to completely separate affordances and perception in order to return to the original Gibsonian notion. In particular, McGrenere and Ho (2000) claim that, according to Gibson, affordances are “independent of the actor’s experience, knowledge, culture, or ability to perceive” (italics added; for a critical analysis of this position see also Bonderup Dohn, 2009). A similar claim is made by Tornvliet (2003): “Gibson labored to make affordances a characteristic of the environment that exists relative to an object but independent of perception.” (italics added). There are reasons to believe that such a strict separation of affordances from perception is not unproblematic.

Independence of perception can be interpreted in three different ways, namely, as independence of: (a) the actor’s general ability to perceive the environment, (b) perceptual information about affordances in ambient energy array, and (c) whether or not the actor, who possesses the general ability to perceive, actually picks up information about an affordance, which information is present in ambient energy array. Arguably, it is only the last interpretation that is both accurate and relevant in the context of Gibson’s theory of affordances.

To claim that affordances are independent of an actor’s general ability to perceive is, apparently, wrong. Gibson’s emphasis on the tight coupling of perception and action implies that actor’s action capabilities include perception. It should be noted that in modern ecological psychology, affordances are commonly defined as “real possibilities for action for a perceiving-acting system” (Wagman and Carello, 2001, emphasis added). That perception is a key factor defining action capabilities can be illustrated with a simple example: if a car driver breaks his or her eyeglasses, the car can become “undrivable”. In that case an object’s affordances change not because something happens to the car but because the driver’s action capabilities become insufficient; and action capabilities become insufficient not because the driver is unable to make physical movements any more, but because of a diminished perceptual function.

The claim that affordances are independent of perceptual information about them in ambient energy array (as, for instance, in the case of a hidden door in a paneled room, see McGrenere and Ho, 2000) is probably formally correct but it is not directly relevant to Gibson's theory of affordances. As already mentioned, Gibson (1979) emphasized that his theory of affordances was predominantly about whether information about affordances is available in ambient light, rather than whether affordances exist or are real. Therefore, Gibson's theory of affordances is specifically concerned with possibilities for action, which are reflected in corresponding structures of ambient arrays of energy and thus can be perceived by the actor, and in the context of the theory it is more or less meaningless to analyze affordances independently of their relation to perceptual information. In this respect, Norman's early perception-centered interpretation of affordances – apart from some terminological problems, as well as certain disagreements about the meaning of "direct pickup" (see Norman, 1988) – is, arguably, generally consistent with the original Gibsonian approach.

But is there a contradiction between Gibson's claims that (a) the theory of affordances is essentially concerned with perceptual information in ambient light and (b) affordances exist even if they are not noticed by the actor? Not really, since information that is present in ambient light may not be actually perceived by the actor. For instance, a pickable mushroom could be unnoticed by a person walking in the woods if the person does not look in the direction of the mushroom.

Therefore, while confusion between affordances and their perception should of course be avoided, a complete separation of affordances from perception would, as argued above, mean going to the opposite, equally undesirable, extreme.

### 44.4.2 Direct and "indirect" perception

Relevance to direct perception appears to be a key factor in the popularity of the concept of affordances in HCI and interaction design. Gaver (1991), points that the main advantage of the ecological perspective is that it "may offer a more succinct approach to the design of artifacts that suggest relevant and desirable actions in an immediate way." (italics added)

One would expect, therefore, that exploring the ways in which direct perception of affordances can be supported with appropriate designs should be a key research issue. However, it has not been the case. The term "Direct Perception" is widely used in HCI literature but analysis of mechanisms, criteria, conditions, and solutions for achieving direct perception of action possibilities of interactive products does not seem to be an actively explored issue in HCI research on affordances.

There are some conceptual obstacles that may have prevented researchers from fruitfully addressing this issue in a concrete and constructive way. On the one hand, Gibson's approach essentially claims that direct perception of our material environment can only be direct. His theory of affordances can be interpreted so that there is no need to support direct perception, since it takes place naturally. It cannot be otherwise: direct perception is the only kind of perception there is.

On the other hand, an opposite argument can be made in case of the visual perception of information expressed in language. Gibson's theory of direct perception does not seem to apply here[13]: apparently, we need to perceive characters comprising a word, and probably look up the word in the dictionary, in order to determine the meaning of the word and associated action possibilities. But in such cases it may appear that perception can only be indirect – so that the task of supporting the transition to direct perception, again, cannot be meaningfully defined.

Therefore, some of the questions, central for putting direct perception of affordances on the agenda of HCI research are: Can the basic principles of ecological

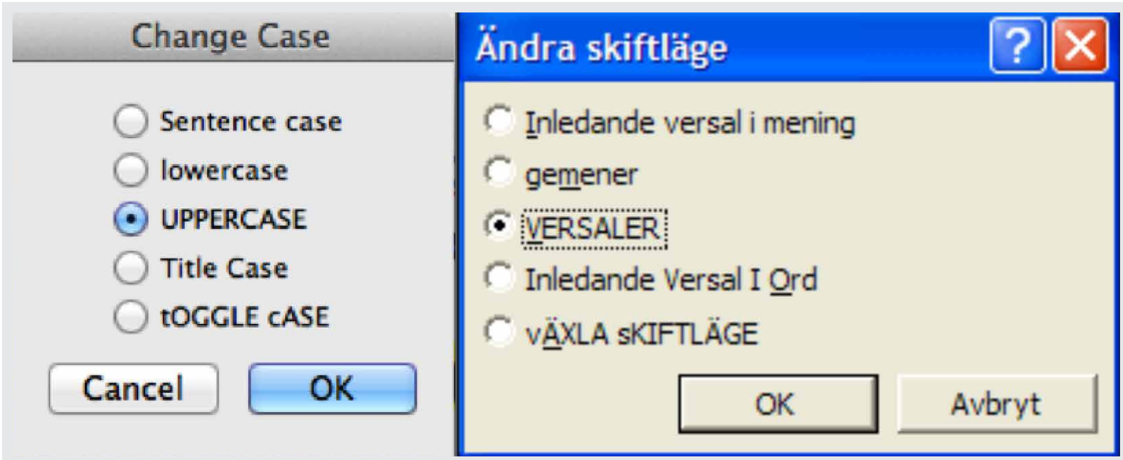
13. For the sake of simplicity, discussions of affordances in Norman's work published between 1989 and 1999 (e.g., Norman, 1993, 1998) are not included

psychology allow for the existence of perception, which is not direct? Can visual language representations be perceived directly? It can be argued that the answer to both of these questions is “yes”.

Eleanor Gibson and Anne Prick (Gibson and Prick 2003), who studied perceptual learning from an ecological perspective, conclude that affordances often need to be discovered, and sometimes it takes much exploration, effort, and patience. Apparently, exploration means that various types of relationship between perceptual information and an affordance are “examined” and “tried out” before the perception of the affordance becomes direct. Therefore, research in ecological psychology suggests that not all perception is direct; direct perception should be considered an accomplishment rather than something that just happens naturally.

At the same time, there is empirical evidence indicating that visual recognition of verbal material can become direct in the sense of visual features being directly used to carry out appropriate actions without language recognition. For instance, evidence obtained in a study of menu selection (Kaptelinin, 1993) suggests that with practice users switch to selecting commands without reading their names, that is, to menu selection based on extracting “non-verbal” visual features, such as screen location or the length of a command name.

How can designers support the transition to direct perception? The general strategy proposed by Still and Dark (2013) is to make designs as consistent as possible. A related, more concrete strategy is to structure ambient optic array so that there is a clear mapping between the structure and appropriate user actions. Consider, for instance, MS Word’s “Change case “dialog box (Figure 44.15). The design of the widget employs certain visual features that make it possible for the user to perceive the widget’s affordances without reading the names of the options. The user does not even need to know the language, as long as the writing system is familiar.




*Copyright status: Unknown (pending investigation). See section "Exceptions" in the copyright terms below.*

FIGURE 44.15: Using visual features to support direct perception: The MS Word "Change Case" dialog box.

## 44.4.3 Culture

Gibson's ecological approach specifically and explicitly deals with perceiving and acting *animals*. The key concepts of the approach, including affordances, are defined in terms of animal-environment interaction. While a variety of illustrating examples, provided by Gibson himself and other proponents of his approach, refer to specifically human objects, such as knives, mailboxes, stairs, airplanes, pictures, and so forth, interaction with these objects is analyzed within the same general framework as interaction of other animals with objects in their respective ecological niches. This perceptive is characteristic for much work in ecological psychology in general. For instance, Eleanor Gibson and Anne Pick (2003) mention an "action instigated by the animal itself, such as driving the truck."

Of course, it is true that we are animals, and this fact has deep implications for how our man-made world is created and experienced. Our built environments, as well as individual things comprising the environments, are as they are to a large extent because we are animals equipped with certain bodies, hands, motor functions, and senses. If we were a different kind of animal, then our houses, cars, airplanes, and computers, if we had them, would look different. Undoubtedly, when designing interactive products, it is important to take into account what ways of action are natural for us as a certain animal species.

However, we humans are also fundamentally unique in a number of respects. As opposed to other animals we are social, cultural creatures: we use language, take part in socially organized collective activities, and employ various artifacts that other animals do not have. Therefore, it is logical to ask: Can an animal-centric theory of affordances provide an account of the *whole* range of human interaction with the world? Can the Gibsonian concept of affordances be used to understand possibilities for specifically *human* action? As mentioned above, these questions have received some attention in HCI research of the last decade.

Some researchers, including Turner (2005), Rizzo (2006) and Vyas et al. (2006), argue that while the framework proposed by Gibson can provide a sensible account of the perception of possibilities for object manipulation and locomotion, that is, immediate interactions with the physical environment, it is difficult to apply the framework to more advanced examples of social, cultural activities. Even some of Gibson's examples, such as using a mailbox, do not easily lend themselves to analysis in terms of layouts, objects, events, and ambient light. While the physical interaction part of using a mailbox is rather straightforward, understanding exactly how people perceive the affordance of sending a letter to a remote location appears to be rather problematic. In general, the analysis of tools by Gibson almost exclusively focuses on simple physical objects, which can in principle be used not only by humans but also by other animals, such as apes. Analysis of more complex tools, which are of main concern to HCI, is virtually missing.

There is growing skepticism in HCI research regarding the potential of the original Gibsonian notion of affordances to serve as a framework for analysis and design of interactive technologies for human use. Considering humans as just another animal species is increasingly perceived as a major limitation of Gibson's theory of affordances in HCI. There are reasons to assume that the general notion of affordances can be fruitfully applied beyond the original Gibsonian scope, that is, animal-environment interaction. Possibilities for human social actions are specified in ambient energy arrays in much the same way as possibilities for physical actions, and they can also be directly perceived. The posture and facial expression of another person may convey an imminent verbal attack as immediately as a view of a cliff would convey a threat of falling off. An open door to a colleague's office may provide as strong a cue to the possibility of striking up an ad hoc conversation as to the possibility for physically going through the doorway. These and similar cases can apparently be described in terms of affordances and their perception, even though the interactions they describe are not limited to object manipulation and locomotion.

Therefore, a key challenge for future research on affordances in HCI appears to be taking into account the context of culture in order to understand how possibilities for *human action* are created, perceived, and can be supported by appropriately designed technology.

### 44.4.4 Affordances of tools

Gibson discusses a variety of tools, such as clubs, knives, and scissors, but he does not systematically explore the issue of what makes tools different from other objects in the environment. For instance, he notes: "A graspable object with a rigid sharp edge affords cutting and scraping (a knife)." (Gibson, 1979). The example suggests that the object's affordances include not only *cut-with-ability* (or scrape-with-ability) but also *graspability*, but the latter is not explicitly considered an affordance. In addition, Gibson describes tools (e.g. scissors) as extensions of the body (e.g., human hand). However, he does not systematically explore how the use of tools affects affordances of other objects in the environment, e.g., how the use of scissors makes a sheet of paper *cuttable*. Therefore, the question, central to HCI, of how affordances of tools are different from affordances of other objects remains open.

Analyses of affordances in HCI do not provide an answer to this question, either. Most of them do not explicitly differentiate between affordances of technological tools and affordances in general (even analyses, which deliberately focus on affordances *of technology*). Take, for instance, Norman's model of action (see Figures 10 and 12 above), employed in several explorations of affordances in HCI. The model does not include an explicit notion of technological tools; it describes how people interact with the "world" and appears to be equally applicable to, say, internet banking and picking berries.

The discussion in Section 3 suggests that some "technology-specific" accounts of affordances can be offered by activity theory and phenomenology. For instance, activity theoretical concept of mediation and phenomenological concepts of breakdowns are explored in, respectively, the mediated action perspective (Kaptelinin and Nardi, 2012) and the analysis of breakdowns (Turner, 2005). However, each of these analyses is currently incomplete and needs to be further developed.

### 44.4.5 Learning

A common assumption about affordances is that perceiving them does not usually require much (or even any) learning; an ability to directly understand affordances is something that we all have. Without any instruction we can see that cliffs afford falling off, small stones afford throwing, and chairs afford sitting. The assumed independence of learning has probably been one of the reasons behind the popularity of affordances among designers. As argued below, however, that assumption is actually a misconception.

To be fair, the misconception is not entirely groundless: in fact, there is virtually no discussion of learning in Gibson's exposition of his theory of affordances. The ability of animals to correctly pick up behaviorally relevant information is, essentially, taken for granted, considered a direct consequence of mutuality between the animal and the environment. On the grand scale of biological evolution the assumption is sound: the very existence (that is, survival) of an animal species testifies that individuals that belong to the species are in principle capable of correctly perceiving affordances of the environment.

However, this argument cannot be directly applied at the level of specific life circumstances of individual animals. When animals are born into the world, their perceptual functions are rudimentary and action capabilities extremely limited. It is only through maturation and practice that they acquire both the ability to act and the ability to pick up information about emerging affordances. Moreover, individual life conditions even for animals of the same species can be very different, so that different affordances are provided to and have to be perceived by the animals. Therefore, for an individual animal the ability to perceive an affordance is not something that can be taken for granted but rather an accomplishment, a result of learning and development.

Studies of perceptual learning and development, conducted within the general framework of Gibson's ecological approach by Eleanor Gibson and her colleagues (e.g. Gibson and Prick, 2003), undoubtedly provide important insights into the centrality of learning in the perception of affordances. A limitation of the studies is that they predominantly deal with processes that take place in *stable* life conditions (e.g., perceptual learning during infancy). In such conditions the outcome of learning is a progressively more advanced adjustment of actors to their environments over extended periods of time. However insightful and important, such studies are of limited relevance to design. New designs are often disruptive. By providing new affordances they may cause significant changes of the environment and create a need for new learning efforts. Anticipating such needs and efficiently supporting users in their learning requires an understanding of how actor-environment mutuality is restored when a disruption takes place - that is, what happens between the moment when new affordances replace old ones and the moment when the actor acquires the ability to directly perceive new affordances. Unfortunately, currently there is a lack of empirical evidence about such phenomena.

It can be concluded, therefore, that explicitly taking affordances into account means that supporting users' discovery of affordances and learning how to use them should be a key designer's concern. Currently there is a lack of evidence on how exactly people learn, unlearn, and re-learn new affordances.

## 44.5 CONCLUSION: REFLECTIONS ON THE PRESENT AND FUTURE OF AFFORDANCES AS AN HCI CONCEPT

### 44.5.1 Interpretation of affordances in different research contexts

As discussed in this chapter, there have been rather dramatic twists and turns in the affordance debate in HCI research since Norman's (Norman 1988) introduction of the concept to the field. Norman's initial interpretation was found to be not entirely consistent with the Gibsonian meaning of the term (Norman, 1999; McGrenere and Ho, 2000; Tornvliet, 2003; Soegaard, 2008). It has been argued that the Gibsonian theory of affordances has a limited relevance to HCI because it does not provide sufficient support for understanding specifically human interaction with – and action through – technology (Albrechtsen et al., 2001; Baerntsen and Trettvik, 2002; Turner, 2005; Rizzo, 2006; Kaptelinin and Nardi, 2012). Repeated attempts to downplay the role of affordances in HCI and interaction design have been made by Norman himself (Norman, 1999, 2008, 2011). Alternative and complementary concepts, such as signifiers and feedforward, have been proposed

(Norman, 2011; Vermeulen et al., 2013). As a result, there is currently a significant degree of uncertainty about the meaning and role of the concept of affordance in the field. While a general understanding of affordances as "action possibilities offered by the environment" is universally accepted, specific interpretations of this general idea are different in different research contexts.

Broadly speaking, the concept of affordances in HCI is used in three related but distinct research agendas, which are predominantly concerned with understanding and supporting, respectively: (a) direct perception, (b) purposeful user action in general, and (c) meaning making. Each of these concerns is associated with a particular perspective on affordances.

Supporting *direct perception* of suitable user actions was the original rationale behind bringing the concept of affordances to HCI (Norman, 1988; Gaver, 1991). The interpretation of affordances in this research agenda is close to the Gibsonian notion, except that "direct perception" is not necessarily understood in the Gibsonian anti-representationalist sense; it can simply mean that no label or instruction is needed to figure out how to use an artifact (Norman, 1988).

Using affordances as an analytical tool to develop *technological support for purposeful human action in general* is an extension of the "direct perception" research agenda. There are two general strategies of using affordances as such an analytical tool. The first strategy is to (a) provide a system of hierarchically organized affordances, that is, action possibilities, which jointly enable the user to attain their meaningful goals and (b) support the user in perceiving these action possibilities (Vicente and Rasmussen, 1990; McGrenere and Ho, 2000). The second strategy is to focus on the "execution-evaluation" cycle of one particular action. The cycle is broken down into specific stages using the model of action, proposed by Norman (1988) and the concept of affordances – alone (Hartson, 2003), or in combination with other related concepts (Vermeulen et al., 2013) – is applied to identify possible ways of supporting the user at each of these stages.

Irrespective of the strategy, perception is playing a key role in the analysis. However, the difference between "direct" and "indirect" perception is usually of secondary importance.

Finally, in a number of relatively recent studies (Turner, 2006; Rizzo, 2006; Vyas et al., 2006; Vyas et al., 2008) it is proposed that the scope of the concept be extended even further, to include *meaning making in social context*. Notions of affordances based on the original Gibsonian concept, are considered limited, as only describing the most basic types of affordances (e.g., "simple affordances", Turner, 2005). It is argued that there is a need for a more advanced notion, according to which affordances are understood as emerging possibilities for individual and collective action in social and cultural contexts, actively constructed by technology users in their everyday practices through both doing and interpretation. The main focus of analysis in this research agenda is not on the "perception – action" cycle but rather on how people generally make sense of the world in terms of action possibilities provided by the environment. Accordingly, perception, as opposed to other research agendas, is either mentioned in passing or not mentioned at all.

Each of these research agendas is associated with its own challenges. Analyses of direct perception of affordances have so far been mostly dealing with physical or physical/virtual actions, such as grasping door handles or clicking on on-screen buttons (e.g., Norman, 1988; Gaver, 1991). Supporting direct perception of possibilities for "non-physical" actions, such as invoking an abstract logical function (see McGrenere and Ho, 2000), while theoretically possible, remains an open issue. The issue is closely related to understanding how direct perception is formed in learning, that is, how an originally indirect process of perception can be transformed into a direct one.

Using affordances as an analytical tool for designing support for purposeful action raises the questions of (a) how the types and properties of affordances, identified in HCI research (e.g., “sequential affordances”, Gaver, 1991, or “degrees of affordances”, McGrenere and Ho, 2000) can be systematically applied in interaction design and (b) whether or not the notion of affordance can be applied to stages of an action rather than whole actions (Hartson, 2003; Vermeulen et al., 2013). Finally, attempts to employ the notion of affordances in studies of meaning making (Turner, 2005; Vyas et al, 2006; Vyas et al., 2008) are yet to provide a clear definition of the new understanding of the term and justify its “added value” compared to other, already existing concepts.

### 44.5.2 Challenges associated with alternative concepts

As argued in the previous section, a number of terminological uncertainties and other conceptual challenges are associated with the concept of affordances. Therefore, a logical question to ask is: Wouldn’t it be a better solution to use instead (at least partly) an alternative or complementary concept proposed in HC research; namely, signifiers or feedforward? Let us consider these alternatives one at a time.

An obvious advantage of the concept of *signifier* (Norman, 2008, Norman 2011) is that it suggests a wide range of possibilities for the designer to orientate, direct, and otherwise support people in their encounters with complex configurations of interactive artifacts, practices, and (social) environments. The designer is encouraged to think about problems and solutions that can be applied to entire everyday contexts rather than narrowly focusing on how to operate a particular device.

The flip side to this advantage, however, is that the meaning of the notion gets extremely broad. Defined as “any perceivable sign for appropriate behavior, whether intentional or unintentional”, a signifier can mean virtually any information available to the senses. Probably the biggest problem caused by the broad meaning of the concept and its strict separation from affordances is that the notion of signifiers provides little guidance in distinguishing good designs from bad ones (even though it is hinted that good designs, as opposed to bad designs, are those that support direct perception, see Norman, 2011).

The concept of feedforward (Vermeulen et al., 2013) also faces the challenge of providing more specific guidance to designers. Such guidance could include, for instance, suggesting specific criteria differentiating more successful designs from less successful ones. A straightforward advice, following from the introduction of the notion of feedforward to design, – that designers should be concerned about informing users about the outcomes of users' actions, – is useful but rather general.

In addition, while there has been significant progress in separating the meaning of "feedforward" from the meaning of "affordance" (Vermeulen et al., 2013), there is still some uncertainty regarding how exactly the concepts can be differentiated from one another. Simply stating that affordances refer to actions while feedforward refers to actions' outcomes, does not seem to be sufficient, since in some cases separating actions from their outcomes may be problematic. A "print preview" seems to be a clear case of feedforward. But does the "close" button of a window inform the user of the actual outcome, a closed window, or about the action of closing (which action may be misapplied, so that the user may accidentally close the wrong window)?

### 44.5.3 Is there a future for affordances as an HCI concept?

What developments in HCI research on affordances can be expected in the future? Which (if any) of the current interpretations of affordances is going to play a central role in the field? Will the term be abandoned in favor of other concepts, such as signifiers or feedforward? While, probably, none can answer these questions with certainty, it would be safe to say that the future of affordances and related concepts in HCI will mostly depend on whether or not they can be clearly defined and shown to be practically relevant.

As argued above, a major problem with current explorations of affordances in HCI is the uncertainty resulting from diverse interpretations of the term in the field. To be a useful conceptual tool, new interpretations of affordances, as well as other proposed concepts, such as signifiers or feedforward, need to be clearly presented and explicitly compared to other interpretations, especially the original Gibsonian meaning, and positioned in a specific research context.

Another important challenge is to make sure a concept is practically relevant and useful, that it provides new insights that help practitioners deal with concrete problems of analysis, design, evaluation, and appropriation of interactive technologies.

When affordance was first proposed as a design concept, it was immediately found practically useful. It suggested, for instance, that making a user interface object look like a familiar physical object can help the user figure out how to operate the object. But this is no longer a new idea: modern interfaces abound with various on-screen buttons, knobs, sliders, and so forth. It appears that the concept of affordances as it was initially introduced to HCI is already well familiar to design practitioners.

Analyses of affordances in HCI research proposed a number of advanced conceptual distinctions, which allow for defining affordances more specifically. Different types and components of affordances can be identified by applying the notions of sequential and nested affordances, degree of affordance, the structure of instrumental affordances, and so forth. These insights open up new possibilities for designers to help people deal with problems associated with modern uses of interactive technologies. Arguably, nowadays users are not particularly puzzled by individual interface objects (e.g., buttons). Instead, they may find it challenging to discover and learn complex configurations of affordances, organized in time and space, assess the effort needed to act out an affordance, and relate mutual affordances of a tool and object of interest to see what action possibilities are offered by the tool. A limitation of advanced theoretical analyses of affordances is that they seldom result in the development of analytical tools suitable for concrete tasks of analysis, design, and evaluation of technology in practical contexts. Operationalizing new theoretical insights in HCI research on affordances is a way to make the research more relevant to practitioners.

In sum, the main challenges for employing new conceptualizations of affordances (or related concepts) in HCI include clarifying the meaning of the concept, as well as its place within a certain research agenda, and making it useful and relevant to designers and other HCI practitioners. Whether or not it can be achieved appears to be critical for determining the future of affordances as an HCI concept.

## 44.6 WHERE TO LEARN MORE

### 44.6.1 Affordances and ecological psychology in general


- Gibson, J. J. The theory of affordances. In: R. Shaw and J. Bransford (eds.) *Perceiving, Acting and Knowing*. Hillsdale, NJ: Erlbaum (1977).
- Gibson, J. J. *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin (1979).
- Gibson, E. J. and Pick, A. D. *An Ecological Approach to Perceptual Learning and Development*. Cary, NC: Oxford University Press (2003).
- Heft, H. *Ecological psychology in context: James Gibson, Roger Barker, and the legacy of William James's radical empiricism*. Lawrence Erlbaum, Mahwah, N.J., (2001).
- Sanders, J.T. An ontology of affordances. *Ecological Psychology*, 9 (1997), 97-112.

- Stoffregen, T. A. Affordances as Properties of the Animal-Environment System. *Ecological Psychology*, 15, 2 (2003), 115-134.
- Wagman, J. and Carello, C. Haptically creating affordances: The user–tool interface. *Journal of Experimental Psychology: Applied*, 9, 4 (2003), 175-186.
- Warren, W. (1995). Constructing an econiche. In J. Flach, P. Hancock, J. Caird, & K. Vicente (Eds.) *Global perspectives on the ecology of human-machine systems*. Hillsdale, NJ: Erlbaum (1995), 210-237.


### 44.6.2 Affordances in HCI


- Baerentsen, K. B. and Trettvik, J. An activity theory approach to affordance. In *Proceedings of NordiCHI 2002*. ACM Press, NY (2002), 51-60.
- Bonderup Dohn, N. Affordances revisited: Articulating a Merleau-Pontian view. *International Journal of Computer-Supported Collaborative Learning,* 4, 2 (2009), 151-170.
- Gaver, W. Technology affordances. In Proceedings of CHI 91. ACM Press: NY (1991), 79-84.
- Gaver, W. The affordances of media spaces for collaboration. In Proceedings of CSCW 92. ACM Press: NY (1992), 17-24.
- Hartson, R. Cognitive, physical, sensory, and functional affordances in interaction design. *Behaviour & Information Technology*, 22, 5 (2003), 315–338.
- Kaptelinin, V. and Nardi, B. Affordances in HCI: Toward a mediated action perspective. In *Proceedings of CHI 2012*. ACM Press, NY (2012).

- McGrenere, J., and Ho, W. Affordances: Clarifying and evolving a concept. In *Proceedings of Graphic Interfaces 2000*. New York: ACM Press (2000), 179-186.
- Norman, D. A. *The Psychology of Everyday Things*. Basic Books, New York (1988).
- Norman, D. A. Affordance, conventions, and design. *interactions*, 6 (3), (1999), 38-43.
- Norman, D. Signifiers, not affordances. *interactions*, 15, 6 (2008).
- Norman, D. *Living with Complexity*. Cambridge, Mass.: The MIT Press (2011).
- Rizzo, A. The origin and design of intentional affordances. In *Proceedings of DIS 2006*. New York: ACM Press (2006).
- Still, J. D., Dark, V. J. Cognitively describing and designing affordances. *Design Studies*, 34 (2013), 285-301.
- Turner, P. Affordance as context. *Interacting with Computers*, 17 (6), (2005), 787-800.
- Vyas, D., Chisalita, C. M., and van der Veer, G. C. Affordance in interaction. In *Proc. ECCE 2006*. Zurich, Switzerland (2006), 92-99.

## 44.7 ACKNOWLEDGMENTS

I would like to thank Mads Soegaard, Bonnie Nardi, Antonio Rizzo, and six anonymous reviewers for their insightful comments on previous versions of the chapter. Peter Larsen's and Rune Arntsen's help with presenting the Holmes stereoscope example is much appreciated.

## 44.8 DISCLAIMER

Some readers may feel the discussion in this chapter is at times not as objective and neutral as one would expect from an encyclopedia chapter. The problem is that current HCI debate on affordances features a number of strong and sometimes conflicting claims, which are difficult to balance with precision. Fortunately, the "comment" functionality of this *online* encyclopedia offers certain affordances for dealing with this issue.

## 44.9 REFERENCES


*Albrechtsen, H., Andersen, Hans H.K., Bodker, S. and Pejtersen, Annelise M. (2001). Affordances in Activity Theory and Cognitive Systems Engineering. RisÃ¸ National Laboratory http://www.risoe.dk/rispubl/SYS/syspdf/ris-r-1287.pdf*

*Beaudouin-Lafon, Michel (2000): Instrumental Interaction: An Interaction Model for Designing Post-WIMP User Interfaces. In: Turner, Thea, Szwillus, Gerd, Czerwinski, Mary, Peterno, Fabio and Pemberton, Steven (eds.) Proceedings of the ACM CHI 2000 Human Factors in Computing Systems Conference April 1-6, 2000, The Hague, The Netherlands. pp. 446-453*

*Bradner, Erin (2001): Social affordances of computer-mediated communication technology: Understanding adoption. In: Proceedings of CHI 01, Extended Abstracts on Human Factors in Computing March 31- April 5, 2001, Seattle, USA. pp. 67-68*

*Brown, John S. and Duguid, Paul (1994): Borderline Issues: Social and Material Aspects of Design. In Human-Computer Interaction, 9 (1) pp. 3-36*

*Bærentsen, Klaus B. (2000): Intuitive User Interfaces. In Scandinavian Journal of Information Systems, 12 pp. 29-60*

*Bærentsen, Klaus B. and Trettvik, Johan (2002): An Activity Theory Approach to Affordance. In: Bertelsen, Olav W., Bødker, Susanne and Kuutti, Kari (eds.) Nordichi 2002 - Proceedings of the Second Nordic Conference on Human-Computer Interaction October 19-23, 2002, Aarhus, Denmark. pp. 51-60*

*Bødker, Susanne (1991): Through the Interface - A Human Activity Approach to User Interface Design. Hillsdale, NJ, Lawrence Erlbaum Associates*

*Bødker, Susanne and Andersen, Peter B. (2005): Complex Mediation. In Human-Computer Interaction, 20 (4) pp. 353-402*

*Cairns, Paul and Thimbleby, Harold (2008): Affordance and symmetry in user interfaces. In The Computer Journal, 51 (6) pp. 650-661*

*Chen, Li-Hao, Lee, Chang-Franw and Tsai, Cheng-Min (2007): Perceiving affordances through perceptual information. In: Poggenpohl, Sharon (ed.) Proceedings of IASDR 07 November 12-15, 2007, Hong Kong, China.*

*Davis, Tehran J., Riley, Michael A., Shockley, Kevin and Cummins-Sebree, Sarah (2010): Perceiving affordances for joint actions. In Perception, 39 (2) pp. 1624-1644*

*Djajadiningrat, Tom, Overbeeke, Kees and Wensveen, Stephan (2002): But how, Donald, tell us how?: on the creation of meaning in interaction design through feedforward and inherent feedback. In: Proceedings of DIS02: Designing Interactive Systems: Processes, Practices, Methods, & Techniques 2002. pp. 285-291*

*Djajadiningrat, Tom, Wensveen, Stephan, Frens, Joep and Overbeeke, Kees (2004): Tangible Products: redressing the balance between appearance and action. In Personal and Ubiquitous Computing, 8 (5) pp. 294-309*

*Dohn, Nina Bonderup (2009): Affordances revisited: Articulating a Merleau-Pontian view. In International Journal of Computer-Supported Collaborative Learning, 4 (2) pp. 151-170*

*Dourish, Paul (2001): Where the Action Is: The Foundations of Embodied Interaction. MIT Press*

*Gaver, William W. (1996): Affordances for interaction: The social is material for design. In Ecological psychology, 8 (2) pp. 111-129*

*Gaver, William W. (1992): The Affordances of Media Spaces for Collaboration. In: Proceedings of the 1992 ACM conference on Computer-supported cooperative work November 01 - 04, 1992, Toronto, Ontario, Canada. pp. 17-24*

*Gaver, William W. (1986): Auditory Icons: Using Sound in Computer Interfaces. In Human-Computer Interaction, 2 (2) pp. 167-177*

*Gaver, William W. (1991): Technology Affordances. In: Robertson, Scott P., Olson, Gary M. and Olson, Judith S. (eds.) Proceedings of the ACM CHI 91 Human Factors in Computing Systems Conference April 28 - June 5, 1991, New Orleans, Louisiana. pp. 79-84*

*Gibson, James J. (1979): The Ecological Approach to Visual Perception. New Jersey, USA, Lawrence Erlbaum Associates*

*Gibson, James J. (1977): The theory of affordances. In: Shaw, Robert and Bransford, John (eds.). "Perceiving, Acting and Knowing". Hillsdale, USA: Lawrence Erlbaum*

*Gibson, Eleanor J. and Pick, Anne D. (2003): An Ecological Approach to Perceptual Learning and Development. Cary, USA, Oxford University Press*

*Giddens, Anthony (1984): The constitution of society: Outline of the theory of structuration. University of California Press*

*Hartson, H. Rex (2003): Cognitive, physical, sensory, and functional affordances in interaction design. In Behaviour and Information Technology, 22 (5) pp. 315-338*

*Heft, Harry (2001): Ecological psychology in context: James Gibson, Roger Barker, and the legacy of William James s radical empiricism. Mahwah, USA, Lawrence Erlbaum*

*Heidegger, Martin (1962): Being and Time. English translation 1962.*

*Ihara, Masayuki, Kobayashi, Minoru and Sakai, Yoshinori (2009): Human affordance. In International Journal of Web Based Communities, 5 (2) pp. 255-272*

*Ilyenkov, Evald V. (1977): The concept of the ideal. In: (ed.). "Philosophy in the USSR: Problems of Dialectical Materialism". Moscow, Russia: Progress Publisherspp. 71-99*

*Judah, Sam (2013). What is skeuomorphism. Retrieved 16 October 2013 from BBC.co.uk: http://www.bbc.co.uk/news/magazine-22840833*

*Kaptelinin, Victor (1993): Item recognition in menu selection: The effect of practice. In: Adjunct proceedings of INTERCHI93 April 24-29, 1993, Amsterdam, Netherlands. pp. 183-184*

*Kaptelinin, Victor and Nardi, Bonnie (2012): Affordances in HCI: toward a mediated action perspective. In: HÃ¶Ã¶k, Kristina (ed.) Proceedings of CHI 2012 May 05-10, 2012, Austin, USA. pp. 967-976*

*Kaptelinin, Victor and Nardi, Bonnie A. (2006): Acting with Technology: Activity Theory and Interaction Design. The MIT Press*

*Kirlik, Alex (2004): On Stoffregen's definition of affordances. In Ecological psychology, 16 (1) pp. 73-77*

*Laarni, Jari, Norros, Leena and Koskinen, Hanna Maria Kaarina (2007): Affordance Table*

- A Collaborative Smart Interface for Process Control. In: Jacko, Julie A. (ed.) HCI International 2007 - 12th International Conference - Part IV 2007. pp. 611-619

Leontyev, Aleksei N. (1978): Activity, consciousness, and personality. Pergamon Press

McGrenere, Joanna and Ho, Wayne (2000): Affordances: Clarifying and Evolving a Concept. In: Proceedings of Graphics Interface 2000 May 15-17, 2000, Montreal, Quebec, Canada. pp. 179-186

Merleau-Ponty, Maurice (1962): Phenomenology of Perception: An Introduction. Routledge

Michaels, Claire F. and Carello, Claudia (1981): Direct Perception. Prentice Hall

Nardi, Bonnie A. (ed.) (1996): Context and Consciousness: Activity Theory and Human-Computer Interaction. MIT Press

Norman, Donald A. (1999): Affordances, Conventions, and Design. In Interactions, 6 (3) pp. 38-41

Norman, Donald A. (2008): Signifiers, not affordances. In Interactions, 15 (6) pp. 18-19

Norman, Donald A. (2010): Living with Complexity. The MIT Press

Norman, Donald A. (1998): The Invisible Computer: Why Good Products Can Fail, the Personal Computer Is So Complex and Information Appliances Are the Solution. MIT Press

Norman, Donald A. (1988): The Psychology of Everyday Things. New York, Basic Books

Norman, Donald A. (1991): Cognitive artifacts. In: Carroll, John M. (ed.). "Designing Interaction: Psychology at the Human-Computer Interface". Cambridge, UK: Cambridge University Presspp. 17-38

Norman, Donald A. (1993): Things That Make Us Smart. Reading, USA, Addison-Wesley

O'Brien, Marita A., Rodgers, Wendy A. and Fisk, Arthur D. (2010). Developing An Organizational Model For Intuitive Design. Georgia Institute of Technology https://smartech.gatech.edu/bitstream/handle/1853/40563/HFA-TR-1001-IntuitiveDesignConceptualOverview.pdf;jsessionid=415E22D63DB9A9E52259BDAE0EC972DF.smart2?sequence=1

Oshlyansky, Lidia, Thimbleby, Harold and Cairns, Paul (2004): Breaking affordance: culture as context. In: Proceedings of the Third Nordic Conference on Human-Computer Interaction October 23-27, 2004, Tampere, Finland. pp. 81-84

Overbeeke, Kees and Hummels, Caroline (2013). Industrial Design. Retrieved 17 February 2014 from [URL to be defined - in press]

*Overbeeke, Kees and Wensveen, Stephan (2003): From perception to experience, from affordances to irresistibles. In: DPPI 2003 - Proceedings of the 2003 International Conference on Designing Pleasurable Products and Interfaces June 23-26, 2003, Pittsburgh, PA, USA. pp. 92-97*

*Raskin, Jef (1994): Viewpoint: Intuitive equals familiar. In Communications of the ACM, 37 (9) pp. 17-18*

*Rasmussen, Jens and Vicente, Kim J. (1989): Coping with Human Errors through System Design: Implications for Ecological Interface Design. In International Journal of Man-Machine Studies, 31 (5) pp. 517-534*

*Rizzo, Antonio (2006): The origin and design of intentional affordances. In: Proceedings of DIS06: Designing Interactive Systems: Processes, Practices, Methods, & Techniques 2006. pp. 239-240*

*Rizzo, A., DeMonte, M., Rubegni, E. and Torsi, S. (2009): The interplay between sensory-motor and intentional affordances. Workshop on Children and Embodied Interaction at IDC09. In: IDC09 June, 2009, Como, Italy.*

*Rizzolatti, Giacomo and Craighero, Laila (2004): The Mirror-Neuron System. In Annual Review of Neuroscience, 27 pp. 169-192*

*Rogers, Yvonne (2012): HCI Theory: Classical, Modern, and Contemporary. In Synthesis Lectures on Human-Centered Informatics, 5 (2) pp. 1-129*

*Rogers, Yvonne (2004): New Theoretical Approaches for HCI. In Annual Review of Information Science and Technology, (38) pp. 1-43*

*Rogers, Yvonne, Sharp, Helen and Preece, Jenny (2011): Interaction Design: Beyond Human - Computer Interaction - third edition. Wiley*

*Sanders, John T. (1997): An ontology of affordances. In Ecological psychology, 9 (1) pp. 97-112*

*Smets, Gerda and Overbeeke, Kees (1994): Industrial design engineering and the theory of direct perception. In Design Studies, 15 (2) pp. 175-184*

*Soegaard, Mads (2013). Affordances. Retrieved 16 October 2013 from Interaction Design Foundation: http://www.interaction-design.org/encyclopedia/affordances.html*

*Souza, Clarisse S. de (2007): Semiotic Engineering of Human-Computer Interaction. Cambridge, USA, The MIT Press*

Still, Jeremiah D. and Dark, Veronica J. (2013): Cognitively describing and designing affordances. In Design Studies, 34 (3) pp. 285-301

Stoffregen, Thomas A. (2003): Affordances as Properties of the Animal-Environment System. In Ecological psychology, 15 (2) pp. 115-134

Suthers, Daniel D. (2006): Technology affordances for intersubjective meaning making: A research agenda for CSCL. In International Journal for Computer-Supported Collaborative Learning, 1 (3) pp. 315-337

Technopedia. Skeuomorphism (n/d). Available at: http://www.techopedia.com/definition/28955/skeuomorphism

Tognazzini, Bruce (2001). The Butterfly Ballot: Anatomy of a Disaster. Retrieved 16 October 2013 from Ask Tog: http://www.asktog.com/columns/042ButterflyBallot.html

Tomasello, Michael (1999): The cultural origins of human cognition. Cambridge, USA, Harvard University Press

Torenvliet, Gerard (2003): We can't afford it!: the devaluation of a usability term. In Interactions, 10 (4) pp. 12-17

Turner, Phil (2005): Affordance as context. In Interacting with Computers, 17 (6) pp. 787-800

Turner, P. and Turner, S. (2002): An Affordance-based Framework for CVE Evaluation. In: Faulkner, Xristine, Finlay, Janet and DÃ©tienne, FranÃ§oise (eds.) Proceedings of the HCI02 Conference on People and Computers XVI September 18-20, 2002, Pisa, Italy. pp. 89-104

Vermeulen, Jo, Luyten, Kris, Hoven, Elise van den and Coninx, Karin (2013): Crossing the bridge over Norman's gulf of execution: Revealing feedforward's true identity. In: Proceedings of HCI 2013 April 27- May 2, 2013, Paris, France. pp. 1931-1940

Vicente, Kim J. and Rasmussen, Jens (1990): The ecology of human-machine systems II: Mediating "direct perception" in complex work domains. In Ecological psychology, 2 (3) pp. 207-249

Vyas, Dhaval, Chisalita, Christina M. and Dix, Alan (2008). Dynamics of Affordances and Implications for Design. University of Twente http://doc.utwente.nl/64769/

Vyas, Dhaval, Chisalita, Cristina M. and Veer, Gerrit C. van der (2006): Affordance in interaction. In: 13th European Conference on Cognitive Ergonomics September 20-22, 2006, Zurich, Switzerland. pp. 92-99

*Wagman, Jeffrey B. and Carello, Claudia (2003): Haptically creating affordances: The user tool interface. In Journal of Experimental Psychology: Applied, 9 (4) pp. 175-186*

*Wagman, Jeffrey B. and Carello, Claudia (2001): Affordances and Inertial Constraints on Tool Use. In Ecological psychology, 13 (3) pp. 173-195*

*Warren, William H. (1995): Constructing an econiche. In: Flach, John M., Hancock, Peter A., Caird, Jeff and Vicente, Kim J. (eds.). "Global perspectives on the ecology of human-machine systems". Hillsdale, USA: Erlbaumpp. 210-237*

*Wikipedia. Preikestolen (n.d.) Available at http://en.wikipedia.org/wiki/Preikestolen*

*Zhang, Ping (2008): Motivational affordances: Fundamental reasons for ICT design and use. In Communications of the ACM, 51 (11)*

*Şahin, Erol, Çakmak, Maya, Doğar, Mehmet R., Uğur, Emre and Üçoluk, Göktürk (2007): To afford or not to afford: A new formalization of affordances toward affordance-based robot control. In Adaptive Behavior - Animals, Animats, Software Agents, Robots, Adaptive Systems, 15 (4) pp. 447-472*

## About the author

### Victor Kaptelinin

© *Victor Kaptelinin*

Has also published under the name of:
"V. Kaptelinin"

Victor Kaptelinin is a Professor at the Department of Information Science and Media Studies, University of Bergen, Norway, and the Department of Informatics, Umeaa University, Sweden. He has held teaching and/or research positions at the Psychological Institute of Russian Academy of Education, Moscow Lomonosov University, and University of California in San Diego, USA. His main research interests are in interaction design, activity theory, and educational use of information technologies.

## Your notes and thoughts on chapter 44

**NOTES:**

# Chapter 52

# Semi-structured qualitative studies

## by Ann Blandford.

Human-Computer Interaction (HCI) addresses problems of interaction design: delivering novel designs, evaluating existing designs, and understanding user needs for future designs. Qualitative methods have an essential role to play in this enterprise, particularly in understanding user needs and behaviours and evaluating situated use of technology. There are, however, a huge number of qualitative methods, often minor variants of each other, and it can seem difficult to choose (or design) an appropriate method for a particular study. The focus of this chapter is on semi-structured qualitative studies, which occupy a space between ethnography and surveys, typically involving observations, interviews and similar methods for data gathering, and methods for analysis based on systematic coding of data. This chapter is pragmatic, focusing on principles for designing, conducting and reporting on a qualitative study and conversely, as a reader, assessing a study. The starting premise is that all studies have a purpose, and that methods need to address the purpose, taking into account practical considerations. The chapter closes with a checklist of questions to consider when designing and reporting studies.

## 52.1 INTRODUCTION

HCI has a focus (the design of interactive systems), but exploits methods from various disciplines. One growing trend is the application of qualitative methods to better understand the use of technology in context. While such methods are well established within the social sciences, their use in HCI is less mature, and there is still controversy and uncertainty about when and how to apply such methods, and how to report the findings (e.g. Crabtree et al., 2009).

This chapter takes a high-level view on how to design, conduct and report semi-structured qualitative studies (SSQSs). Its perspective is complementary to most existing resources (e.g. Adams et al., 2008; Charmaz, 2006; Lazar et al., 2010; Smith, 2008; onlineqda.hud.ac.uk), which focus on method and principles rather than basic practicalities. Because 'method' is not a particularly trendy topic in HCI, I draw on the methods literature from psychology and the social sciences as well as HCI. Rather than starting with a particular method and how to apply it, I start from the purpose of a study and the practical resources and constraints within which the study must be conducted.

I do not subscribe to the view that there is a single right way to conduct any study: that there is a minimum or maximum number of participants; that there is only one way to gather or analyse data; or that validation has to be achieved in a particular way. As Willig (Willig, 2008, p.22) notes, "Strictly speaking, there are no 'right' or 'wrong' methods. Rather, methods of data collection and analysis can be more or less appropriate to our research question." Woolrych et al. (2011) draw an analogy with ingredients and recipes: the art of conducting an effective study is in pulling together appropriate ingredients to construct a recipe that is right for the occasion – i.e., addresses the purpose of the study while working with available resources.

The aim of this chapter is to present an overview of how to design, conduct and report on SSQSs. The chapter reviews methodological literature from HCI and the social and life sciences, and also draws on lessons learnt through the design, conduct and reporting of various SSQSs. The chapter does not present any method in detail, but presents a way of thinking about SSQSs in order to study users' needs and situated practices with interactive technologies.

The basic premise is that, starting with the purpose of a study, the challenge is to work with the available resources to complete the best possible study, and to report it in such a way that its strengths and limitations can be inspected, so that others can build on it appropriately. The chapter summarises and provides pointers to literature that can help in research, and at the end a checklist of questions to consider when designing, conducting, reporting on, and reviewing SSQSs. The aim is to deliver a reference text for HCI researchers planning semi-structured qualitative studies.

### 52.1.1 What is an SSQS?

The term 'semi-structured qualitative study' (SSQS) is used here to refer to qualitative approaches, typically involving interviews and observations, that have some explicit structure to them, in terms of theory or method, but are not completely structured. Such studies typically involve systematic, iterative coding of verbal data, often supplemented by data in other modalities.

Some such methods are positivist, assuming an independent reality that can be investigated and agreed upon by multiple researchers; others are constructivist, or interpretivist, assuming that reality is not 'out there', but is constructed through the interpretations of researchers, study participants, and even readers.

In the former case, it is important that agreement between researchers can be achieved. In the latter case, it is important that others are able to inspect the methods and interpretations so that they can comprehend the journey from an initial question to a conclusion, assess its validity and generalizability, and build on the research in an informed way.

In this chapter, we focus on SSQSs addressing exploratory, open-ended questions, rather than qualitative data that is incorporated into hypothetico-deductive research designs. Kidder and Fine (Kidder and Fine 1987, p.59) define the former as “big Q” and the latter as “small q”, where “big Q” refers to “unstructured research, inductive work, hypothesis generation, and the development of ‘grounded theory’”. Their big Q encompasses ethnography (Section 52.1.4) as well as the SSQSs that are the focus here; the important point is that SSQSs focus on addressing questions rather than testing hypotheses: they are concerned with developing understanding in an exploratory way.

One challenge with qualitative research methods in HCI is that there are many possible variants of them and few names to describe them. If every one were to be classed as a ‘method’ there would be an infinite number of methods. However, starting with named methods leaves many holes in the space of possible approaches to data gathering and analysis. There are many potential methods that have no name and appear in no textbooks, and yet are potentially valid and valuable for addressing HCI problems.

This contrasts with quantitative research. Within quantitative research traditions – exemplified by, but not limited to, controlled experiments – there are well-established ways of describing the research method, such that a suitably knowledgeable reader can assess the validity of the claims being made with reasonable certainty, for example, hypothesis, independent variable, dependent variable, power of test, choice of statistical test, number of participants.

The same is not true for SSQSs, where there is no hypothesis – though usually there is a question, or research problem – where the themes that emerge from the data may be very different from what the researcher expected, and where the individual personalities of participants and their situations can have a huge influence over the progress of the study and the findings.

Because of the shortage of names for qualitative research methods, there is a temptation to call a study an 'ethnography' or a 'Grounded Theory' (both described below: Section 52.1.4 and Section 52.1.5) whether or not they have the hallmarks of those methods as presented in the literature. Data gathering for SSQSs typically involves the use of a semi-structured interview script or a partial plan for what to focus attention on in an observational study.

There is also some structure to the process of analysis, including systematic coding of the data, but usually not a rigid structure that constrains interpretation, as discussed in Section 52.7. SSQSs are less structured than, for example, a survey, which would typically allow people to select from a range of pre-determined possible answers or to enter free-form text into a size-limited text box. Conversely, they are more structured than ethnography – at least when that term is used in its classical sense; see Section 52.1.4.

### 52.1.2 A starting point: problems or opportunities

Most methods texts (e.g. Cairns and Cox, 2008; Lazar et al., 2010; Smith, 2008; Willig, 2008) start with methods and what they are good for, rather than starting with problems and how to select and adapt research methods to address those problems. Willig (Willig, 2008, p.12) even structures her text around questions about each of the approaches she presents:

> *"What kind of knowledge does the methodology aim to produce? … What kinds of assumptions does the methodology make about the world? … How does the methodology conceptualise the role of the researcher in the research process?"*

If applying a particular named method, it is important to understand it in these terms to be able to make an informed choice between methods. However, by starting at the other end – the purpose of the study and what resources are available – it should be possible to put together a suitable plan for conducting a SSQS that addresses the purpose, makes relevant assumptions about the world, and defines a suitable role for the researcher.

Some researchers become experts in particular methods and then seek out problems that are amenable to that method; for example, drawing from the social sciences rather than HCI, Giorgi and Giorgi (Giorgi and Giorgi, 2008) report seeking out research problems that are amenable to their phenomenology approach. On the one hand, this enables researchers to gain expertise and authority in relation to particular methods; on the other, this risks seeing all problems one way: "To the man who only has a hammer, everything he encounters begins to look like a nail", to quote Abraham Maslow.

HCI is generally problem-focused, delivering technological solutions to identified user needs. Within this, there are two obvious roles for SSQSs: understanding current needs and practices, and evaluating the effects of new technologies in practice. The typical interest is in how to understand the 'real world' in terms that are useful for interaction design. This can often demand a '*bricolage*' approach to research, adopting and adapting methods to fit the constraints of a particular problem situation. On the one hand this makes it possible to address the most pressing problems or questions; on the other, the researcher is continually having to learn new skills, and can always feel like an amateur.

In the next section, I present a brief overview of relevant background work to set the context, focusing on qualitative methods and their application in HCI. Subsequent sections cover an approach to planning SSQSs based on the PRET A Rapporter framework (Blandford et al., 2008) and discuss specific issues including the role of theory in SSQSs, assessing and ensuring quality in studies, and various roles the researcher can play in studies. This chapter closes with a checklist of issues to consider in planning, conducting and reporting on SSQSs.

### 52.1.3 A brief overview of qualitative methods

There has been a growing interest in the application of qualitative methods in HCI. Suchman's (Suchman, 1987) study of situated action was an early landmark in recognising the importance of studying interactions in their natural context, and how such studies could complement the findings of laboratory studies, whether controlled or employing richer but less structured techniques such as think aloud.

Sanderson and Fisher (Sanderson and Fisher, 1994) brought together a collection of papers presenting complementary approaches to the analysis of sequential data (e.g., sequences of events), based on a workshop at CHI 1992. Their focus was on data where sequential integrity had been preserved, and where sense was made of the data through relevant techniques such as task analysis, video analysis, or conversation analysis. The interest in this collection of papers is not in the detail, but in the recognition that semi-structured qualitative studies had an established place in HCI at a time when cognitive and experimental methods held sway.

Since then, a range of methods have been developed for studying people's situated use and experiences of technology, based around ethnography, diaries, interviews, and similar forms of verbal and observable qualitative data (e.g. Lindtner et al. 2011; Mackay 1999; Odom et al. 2010; Skeels & Grudin 2009).

Some researchers have taken a strong position on the appropriateness or otherwise of particular methods. A couple of widely documented disagreements are briefly discussed below. This chapter avoids engaging in such 'methods wars'. Instead, the position, like that of Willig (2008) and Woolrych et al. (2011), is that there is no single correct 'method', or right way to apply a method: the textbook methods lay out a space of possible ways to conduct a study, and the details of any particular study need to be designed in a way that maximises the value, given the constraints and resources available. Before expanding on that theme, we briefly review ethnography – as applied in HCI – and Grounded Theory, as a descriptor that is widely used to describe exploratory qualitative studies.

### 52.1.4 Ethnography: the all-encompassing field method?

Miles and Huberman (Miles and Huberman 1994, p.1) suggest, "The terms ethnography, field methods, qualitative inquiry, participant observation, … have become practically synonymous". Some researchers in HCI seem to treat these terms as synonymous too, whereas others have a particular view of what constitutes 'ethnography'. For the purposes of this chapter, an ethnography involves observation of technology-based work leading to rich descriptions of that work, without either the observation or the subsequent description being constrained by any particular structuring constructs. This is consistent with the view of Anderson (1994), and Randall and Rouncefield (2013).

Crabtree et al. (2009) present an overview – albeit couched in somewhat confrontational terms – of different approaches to ethnography in HCI. Button and Sharrock (2009) argue, on the basis of their own experience, that the study of work should involve "ethnomethodologically informed ethnography", although they do not define this succinctly. Crabtree et al. (Crabtree et al. 2000 , p.666) define it as study in which “members’ reasoning and methods for accomplishing situations becomes the topic of enquiry”.

Button and Sharrock (2009) present five maxims for conducting ethnomethodological studies of work: keep close to the work; examine the correspondence between work and the scheme of work; look for troubles great and small; take the lead from those who know the work; and identify where the work is done. They emphasise the importance of paying attention, not jumping to conclusions, valuing observation over verbal report, and keeping comprehensive notes. However, their guidance does not extend to any form of data analysis. In common with others (e.g. Heath & Luff, 1991; Von Lehn & Heath, 2005), the moves that the researcher makes between observation in the situation of interest and the reporting of findings remain undocumented, and hence unavailable to the interested or critical reader.

According to Randall and Rouncefield (2013), ethnography is “a qualitative orientation to research that emphasises the detailed observation of people in naturally occurring settings”. They assert that ethnography is not a method at all, but that data gathering “will be dictated not by strategic methodological considerations, but by the flow of activity within the social setting”.

Anderson (1994) emphasises the role of the ethnographer as someone with an interpretive eye delivering an account of patterns observed, arguing that not all fieldwork is ethnography and that not everyone can be an ethnographer. In SSQSs, our focus is on methods where data gathering and analysis are more structured and open to scrutiny than these flavours of ethnography.

### 52.1.5 Grounded Theory: the SQSS method of choice?

I am introducing Grounded Theory (GT) early in this chapter because the term is widely used as a label for any method that involves systematic coding of data, regardless of the details of the study design, and because it is probably the most widely applied SSQS method in HCI.

GT is not a theory, but an approach to theory development – grounded in data – that has emerged from the social sciences. There are several accounts of GT and how to apply it, including Glaser and Strauss (2009), Corbin and Strauss (2008), Charmaz (2006), Adams et al. (2008), and Lazar et al. (2010).

Historically, there have been disputes on the details of how to conduct a GT: the disagreement between Glaser and Strauss, following their early joint work on Grounded Theory (Glaser and Strauss, 2009), has been well documented (e.g. Charmaz, 2008; Furniss et al., 2011a, Willig, 2008). Charmaz (2006) presents an overview of the evolution of different strains of GT prior to that date.

Grbich (2013) identifies three main versions of GT, which she refers to as Straussian, involving a detailed three-stage coding process; Glaserian, involving less coding but more shifting between levels of analysis to relate the details to the big picture; and Charmaz's, which has a stronger constructivist emphasis.

Charmaz (Charmaz 2008, p.83) summarises the distinguishing characteristics of GT methods as being:

- Simultaneous involvement in data collection and analysis;
- Developing analytic codes and categories "bottom up" from the data, rather than from preconceived hypotheses;
- Constructing mid-range theories of behaviour and processes;
- Creating analytic notes, or memos, to explain categories;
- Constantly comparing data with data, data with concept, and concept with concept;
- Theoretical sampling – that is, recruiting participants to help with theory construction by checking and refining conceptual categories, not for representativeness of a given population;
- Delaying literature review until after forming the analysis.

There is widespread agreement amongst those who describe how to apply GT that it should include interleaving between data gathering and analysis, that theoretical sampling should be employed, and that theory should be constructed from data through a process of constant comparative analysis. These characteristics define a region in the space of possible SSQSs, and highlight some of the dimensions on which qualitative studies can vary. I take the position that the term 'Grounded Theory' should be reserved for methods that have these characteristics, but even then it is not sufficient to describe the method simply as a Grounded Theory without also presenting details on what was actually done in data gathering and analysis.

As noted above, much qualitative research in HCI is presented as being Grounded Theory, or a variant on GT. For example, Wong and Blandford (2002) present Emergent Themes Analysis as being "based on Grounded Theory but tailored to take advantage of the exploratory and efficient data collection features of the CDM" – where CDM is the Critical Decision Method (Klein et al., 1989) as outlined in Section 52.6.4.

McKechnie et al. (2012) describe their analysis of documents as a Grounded Theory, and also discuss the use of inter-rater reliability – both activities that are inconsistent with the distinguishing characteristics of GT methods if those are taken to include interleaving of data gathering and analysis and a constructivist stance. GT has been used as a 'bumper sticker' to describe a wide range of qualitative analysis approaches, many of which diverge significantly from GT as presented by the originators of that technique and their intellectual descendants.

Furniss et al. (2011a) present a reflective account of the experience of applying GT within a three-year project, focusing particularly on pragmatic 'lessons learnt'. These include practical issues such as managing time and the challenges of recruiting participants, and also theoretical issues such as reflecting on the role of existing theory – and the background of the analyst – in informing the analysis.

Being fully aware of relevant existing theory can pose a challenge to the researcher, particularly if the advice to delay literature review is heeded. If the researcher has limited awareness of relevant prior research in the particular domain, it can mean 'rediscovery' of theories or principles that are, in fact, already widely recognized, leading to the further question, "So what is new?" We return to the challenge of how to relate findings to pre-existing theory, or literature that emerges as being important through the analysis, in Section 52.9.1.

## 52.2 PLANNING AND CONDUCTING A STUDY: PRET A RAPPORTER

Research generally has some kind of objective (or purpose) and some structure. A defining characteristic of SSQSs is that they have shape… but not too much: that there is some structure to guide the researcher in how to organise a study, what data to gather, how to analyse it, etc., but that that structure is not immutable, and can adapt to circumstances, evolving as needed to meet the overall goals of the study. The plan should be clear, but is likely to evolve over the course of a study, as understanding and circumstances change.

Thomas Green used to remind PhD students to "look after your GOST", where a GOST is a Grand Overall Scheme of Things – his point being that it is all too easy to let the aims of a research project and the fine details get out of synch, and that they need to be regularly reviewed and brought back into alignment.

We structure the core of this chapter in terms of the PRET A Rapporter (PRETAR) framework (Blandford et al., 2008a), a basic structure for designing, conducting and reporting studies. Before presenting this structure, though, it is important to emphasise the basic interconnectedness of all things: in the UK a few years ago there was a billboard advertisement, "You are not stuck in traffic. You are traffic" (Figure 52.1).

It is impossible to separate the components of a study and treat them completely independently – although they have some degree of independence. The style of data gathering influences what analysis can be performed; the relationship established with early participants may influence the recruitment of later participants; ethical considerations may influence what kinds of data can be gathered, etc. We return to this topic of interdependencies later; first, for simplicity of exposition, we present key considerations in planning a study using the PRETAR framework.




Figure 52.1: An example of interconnectedness.

The PRETAR framework draws its inspiration from the DECIDE framework proposed by Rogers et al. (2011), but has a greater emphasis on the later – analysis and reporting – stages that are essential to any SSQS:

- Purpose: every study has a purpose, which may be more or less precisely defined; methods should be selected to address the purpose of the study. The purpose of a study may change as understanding develops, but few people are able to conduct an effective study without some idea of why they are doing it.
- Resources and constraints: every study has to be conducted with the available resources, also taking account of existing constraints that may limit what is possible.
- Ethical considerations often shape what is possible, particularly in terms of how data can be gathered and results reported.
- Techniques for data gathering need to be determined (working with the available resources to address the purpose of the study).
- Analysis techniques need to be appropriate to the data and the purpose of the study.
- Reporting needs to address the purpose of the study, and communicate it effectively to the intended audiences. In some cases, this will include an account of how and why the purpose has evolved, as well as the methods, results, etc.

Some authors have focused attention on one of these steps; for example, Kvale and Brinkmann (2009) focus primarily on data gathering while Miles and Huberman (1994), Grbich (2013) and Braun and Clarke (2006) focus on analysis, and Morse (1997) and Wolcott (2009) on reporting and other aspects of closing off a research project. However, these steps are not independent, and are typically interleaved in SSQSs. What matters is that they remain coherent – that there is a clear GOST. For example, the researcher's view of the purpose of a study may evolve as their understanding matures through data gathering and analysis.

As noted above (Section 52.1.5), GT is based on tight coupling between data gathering and analysis; other analysis techniques assume no such coupling. These steps provide a useful structure to organise our discussion on planning and conducting a study, but should not be regarded as strictly sequential or independent.

## 52.3 PURPOSE

Every study has a purpose. That purpose might be to better understand:

- 'work', broadly conceived, and how interactive technologies support or fail to support that work (e.g. Hartswood et al., 2003; Hughes et al., 1994);
- people's experiences with a particular kind of technology (e.g. Palen, 1999; Kindberg et al., 2005; Mentis et al., 2013);
- how people exploit technologies to support cognition (e.g. Hutchins 1995);
- how people make sense of information with and without particular technological support (e.g. Attfield and Blandford, 2011); or
- many other aspects of the design and use of interactive technologies.

The recent 'turn to the wild' (Rogers, 2012), in which novel products are designed in situ, working directly with the intended users of those products, introduces yet more possible purposes for qualitative studies: to understand how a new product changes attitudes and behaviours, and how the design of the product might be adapted to better support people's needs and aspirations.

Crabtree et al. (2009)argue that the purpose of an ethnographic study in HCI is to inform system design. They claim that ethnographic research to inform systems design has shifted from a study of work towards a study of culture, and that this shift is “harmful”. In creating an either/or stand-off, the authors pit the contrasting ethnographic focuses against each other, apparently disregarding the possibility that each has a place in informing design, and that different ethnographic studies of the same context can serve different purposes. The same is true of SSQSs: they can address many different questions that inform design – though whether ‘informing design’ should mean that there are explicitly stated ‘implications for design’ is a further question.

Crabtree et al. (2009) suggests that there is a widespread expectation that studies will always include “implications for design” as an explicit theme, and that this is expected in reporting even if that was not the purpose of the study. He argues that designers need a rich understanding of the situation for which they are designing, and that one of the important roles for ethnography is to expose and describe that cultural context for design, without necessarily making the explicit link to implications for design. This might be regarded as an argument for helping designers to put themselves in the user’s shoes when they are not designing for themselves, for example, designing specialist products to support work or other activities in which they are not experts themselves.

Obviously, not all studies should be qualitative, and certainly not all should be semi-structured. Yardley (Yardley 2000, p.220) differentiates between the typical purposes of qualitative and quantitative studies:

*“Quantitative studies … ensure the ‘horizontal generalization’ of their findings across research settings … qualitative researchers aspire*

*instead to … 'vertical generalization', i.e., an endeavour to link the particular to the abstract and to the work of others".*

. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .

In HCI, qualitative studies – whether structured, semi-structured or ethnographic – most typically focus on understanding technology use, or future technology needs, in situated settings, recognizing that laboratory studies are limited when it comes to investigating issues around real-world use.

In summary, there are many purposes for which SSQSs are well suited. There are others that demand other techniques, such as controlled laboratory studies or ethnography (in the sense discussed in Section 52.1.4); what matters is that the study design suits the study purpose.

There are, of course, purposes that cannot, in practice, be addressed reliably, for legal, safety, privacy, ethical or similar reasons. For example, in safety-critical situations, the presence of researchers could be a distraction when conditions become demanding, so it may not be possible to study the details of interactions at times of greatest stress. It is not theoretically or practically possible to address every imaginable research question in HCI.

The centrality of purpose is emphasised by Wolcott (Wolcott 2009, p.34), who advocates "a candidate for the opening sentence for scholarly writing: 'The purpose of this study is…'". While the purpose should drive the study design, and might evolve in the light of early study findings, it may also have to be crafted to fit the available resources.

## 52.4 RESOURCES AND CONSTRAINTS

Every study has to be designed to work with the available resources. Where resources are limited by, for example, time or budgetary restraints, it is necessary to 'cut your coat according to your cloth' – i.e., to fit ambitions and hence purpose to what is possible with the available resources.

Resource considerations need to cover – at least – time, funding, equipment available for data collection and analysis, availability of places to conduct the study, availability of participants, and expertise. In many cases, it is also necessary to have advocacy and support from people with influence in the intended study settings. Of these resources, three that merit further discussion are advocacy, participants and expertise.

### 52.4.1 Advocacy

Sometimes, studies are devised and run in collaboration with 'problem owners' (e.g. Randell et al, 2013), but other studies are conceived by a research team outside a particular domain setting. In some cases, it is essential to get support from a domain specialist.

For example, in the work of the author and co-workers, with a shift in emphasis in healthcare from hospital to home, we are interested in how medical devices are taken up and used in the home, and how products that were originally developed for use by clinical staff in hospitals can be adapted to be suitable for home use. There are some products that are well established for home use, such as nebulisers, blood glucose monitors and dialysis machines, and others that are making the transition from hospital to home, such as patient-controlled analgesia and intravenous administration of chemotherapy.

We followed several lines of enquiry to identify clinicians who expressed an interest in patients' experiences of intravenous therapies at home, but all eventually drew a blank. In contrast, we identified several nephrologists who were sufficiently interested in patients' experience of home haemodialysis to introduce us to their patients. This has led to a productive study (e.g. Rajkomar et al., 2013).

In other cases, it may be important to obtain permission to conduct a study in a particular location; for example, Perera (2006)investigated under what circumstances people forgot their chip-and-pin cards in shops, and she needed to obtain permission from shop managers to conduct observational studies on their premises.

In another case, in work for a Masters project (O'Connor, 2011), support was needed from the nursing manager in the emergency department; as soon as she was contacted, it was clear that she was keen to support others in their education, and this was the main factor in her supporting the project, rather that its inherent interest. In yet other cases, there is no particular advocate or manager to involve; for example, studies of diabetes patients (e.g. O'Kane & Mentis, 2012) involved direct recruitment of participants without mediation from specialists.

There may be a hidden cost in negotiating support from advocates, but this often brings with it the benefits of close engagement with the study domain, introductions to potential participants, and longer-term impact through the engagement of stakeholders.

### 52.4.2 Participants: recruitment and sampling

When recruiting participants for a study, with or without the advocacy of an intermediary, it is important to consider their motivations for participation. This is partly coupled with ethical considerations (Section 52.5), and partly with how to incentivise people to participate at all. People may agree or elect to participate in studies for many different reasons: maybe it is low-cost (in terms of time and effort), and people just want to be helpful.

This was probably the case in the studies of ambulance control carried out by Blandford & Wong (2004). The immediate benefits to participants, ambulance controllers, were relatively small, beyond the sense that someone else was interested in their work and valued their expertise. However, the costs of participation were also low – continue to do your job as normal, and talk about the work in slack periods when you would otherwise simply be waiting for the next call.

In other cases, participants may be inherently interested in the project – as in some of our work on serendipity (Makri & Blandford, 2012) – or perceive some personal benefit, for example, in reflecting on how you manage your time (Kamsin et al., 2012). Some people may participate for financial reward, or to return a favour. And of course, people's motivations for participating may be mixed.

Where the topic is one that participants might be sensitive about, for example, intimate health issues, it can sometimes help to have pre-existing common ground between the researcher doing data gathering and the participant, such as being of the same sex or a similar age. Where multiple researchers are available, this might mean matching them well to participants; where there is a single researcher, it might mean reviewing the purpose of the study to be sure that data gathering is likely to be productive. In the section on ethics (Section 52.5), we discuss relationships with participants and how these relate to recruitment and motivations for participating.

The choice of approaches to recruitment depends on the purpose of the study and the kinds of participant needed. Possible approaches include:

- Direct contact – e.g. approaching individuals in the workplace, with authorisation from local managers if needed, or approaching people in public spaces, with due regard for safety, informed consent, etc..
- Mediated contact: an introduction by someone else, such as a line manager in the workplace, another 'gatekeeper' – for example, a teacher, or the organiser of a relevant 'special interest' group – or friends or other participants.

- Advertising: on noticeboards in physical space, through targeted email lists, via online lists and social networks.

As social media and other technologies evolve, new approaches to recruiting study participants are emerging. What matters is that the approach to recruitment is effective in terms of recruiting both a suitable number of participants and appropriate participants for the aims of the study.

Two questions that come up frequently are how many participants should be included and how they should be sampled. The answer to both is ‘it depends’ – on the aims of the study, and what is possible with the available resources.

Although not common in HCI, it is possible to conduct a study with a single participant, as a rich case study. For example, Attfield et al. (2008) gathered observations, interview data and examples of artefacts produced from a single journalist as that journalist prepared an article from inception to publication. The aim of the study was to understand the phases of work, how information was transformed through that work, and how technology supported the work.

Such a case study provides a rich understanding of the interaction, but care has to be taken over generalizing: ideally, such a case will be compared with known features of comparable cases, in terms of both similarities and contrasts. In poorly understood areas, even a single rich case study can add to our overall understanding of the design, deployment and use of interactive technologies. But most studies involve many more participants than one.

Smith (Smith 2008, p.14) draws the distinction between idiographic and nomothetic research as follows:

*“The nomothetic approach assumes that the behaviour of a particular person is the outcome of laws that apply to all, and the aim of science*

*is to reveal these general laws. The idiographic approach would, in contrast, focus on the interplay of factors which may be quite specific to the individual.”*

In other words, nomothetic research relies more on large samples and statistical techniques to establish generalizations, for example, through controlled experiments. SSQSs typically involve much smaller numbers of participants, occasionally as few as one, but more commonly 10-20, but gathers rich data with each. In this sense, SSQSs are idiographic, and care must be taken with generalizing beyond the study setting. This is a topic to which we return in Section 52.10.1.

GT researchers resist specifying numbers of participants required. Rather, they advocate continuing to gather data until the theoretical categories of the analysis are saturated. Charmaz (Charmaz 2006, p. 113) explains: “Categories are ‘saturated’ when gathering fresh data no longer sparks new theoretical insights, nor reveals new properties of your core theoretical categories”. In other words, you stop gathering data when it no longer advances the study. This presupposes an iterative approach to data gathering and analysis, which is the case for GT, but not for other styles of qualitative research where all data may be gathered before analysis commences.

In practice, there are often pragmatic factors that determine how many participants to involve in a study. One might be the time available: it can take a long time to recruit each participant, to arrange and conduct data gathering, and analyse the data. Another might be the availability of participants who satisfy the recruitment criteria, for example, performing a particular role in an organisation or having particular experience. A shorter study with fewer participants needs to be more focused in order to deliver insight, because otherwise it risks delivering shallow data from which it is almost impossible to derive valuable insight.

Thought must be given to how to sample participants. Sometimes, the criteria are quite broad , for example, people who enjoy playing video games, and it is possible to recruit through public advertising. Sometimes, they are focused, such as on people with a particular job role within an organisation.

For other studies, the aim might be to obtain a representative sample; for example, in a study of lawyers' use of information resources (Makri et al., 2008), our aim was to involve lawyers across the range of seniority, from undergraduate students to senior partners in a law firm and professors in a university law department.

However, in qualitative research it is rare to aim for probability sampling, as one would for quantitative studies. Marshall (1996) discusses three different approaches to sampling for qualitative research: convenience, judgement (also called purposeful), and theoretical.

- Convenience sampling involves working with the most accessible participants, and is therefore the easiest approach.
- Judgment sampling, in which the "researcher actively selects the most productive sample to answer the research question" (p. 523), is the most commonly used in HCI.
- Theoretical sampling is advocated within GT, and involves recruiting participants who are most likely to help build the theory that is emerging through data gathering and analysis.

Miles and Huberman (Miles and Huberman 1994, p.28) list no fewer than 16 different approaches to sampling, such as maximum variation, extreme or deviant case, typical case, and stratified purposeful, each with a particular value in terms of data gathering and analysis.

An approach to sampling that can be particularly useful for accessing hard-to-reach populations, for example, people using a particular specialist device, is snowball sampling, where each participant introduces the researcher to further participants who satisfy their inclusion criteria.

Atkinson and Flint (2001) highlight some of the limitations of this approach, in terms of participant diversity and consequent generalizability of findings. Slightly tongue-in-cheek, they also describe “scrounging sampling”: the increasingly desperate acquisition of participants to make up numbers almost regardless of suitability. While few authors are likely to admit to applying scrounging sampling as a strategy for recruitment, it is important to explain clearly how and why participants have been recruited and the likely consequences of the recruitment strategy on findings.

The same issue arises when there might be barriers to recruitment. Buckley et al. (2007) highlight the dangers of ‘consent bias’, whereby those with more positive outcomes are more likely to agree to participate in a study. Although most of the literature on consent bias relates to healthcare studies, there are similar risks in HCI studies, particularly where the technology under investigation is related to a sensitive personal issue, such as behaviour change technologies. Atkinson and Flint (2001) also highlight the risks of ‘gatekeeper bias’, where those in authority – for example, clinicians or teachers – filter out potential participants whom they consider less suitable.

In summary, when planning a study, it is important to consider questions of recruitment and relationship management:

- Who the appropriate participants are and how they should be recruited;
- Where and when to work with them in data gathering; and
- How (or whether) to engage with them more broadly from the start to the end of a study.

Throughout recruitment, study design and data analysis, it is important to remain aware of participants' motivations for participating, and their expectations of the outcome, whether this is, for example, the expectation of novel interaction designs, or simply to gain the experience of participating. These factors are addressed further in Section 52.5.

When dealing with sensitive topics where people may have reasons for sharing or withholding certain information, or for behaving in particular ways, it is also important to be aware of motivations and their possible effects on the data that is gathered. Such considerations imply the need (a) to review data gathering techniques to maximise the likelihood of gathering valid data (see Section 52.6) and (b) to reflect on the data quality and implications for the findings (see Section 52.10.1).

### 52.4.3 Expertise of the research team

There are at least two aspects to expertise: that in qualitative research and that in the study domain.

There is no shortcut to acquiring expertise in qualitative research. Courses, textbooks and research papers provide essential foundations, and different resources resonate with – and are therefore most useful to – different people. Corbin and Strauss (Corbin and Strauss 2008, p.27) emphasise the importance of planning and practice:

"Persons sometimes think that they can go out into the field and conduct interviews or observations with no training or preparation. Often these persons are disappointed when their participants are less than informative and the data are sparse, at best."

Kidder and Fine (Kidder and Fine 1987, p.60) describe the evolving focus of qualitative research: "A daily chore of a participant observer is deciding which question to ask next of whom." There is no substitute for planning, practice and reflecting on what can be learnt from each interview or observation session.

Yardley(Yardley 2000, p.218), comments on the trend towards precisely defined methods:

> *"This trend is fuelled by the tendency of those who are new to qualitative research, and dismayed by the scope and complexity of the field, to adhere gratefully to any set of clear-cut procedures provided by proponents of a particular form of analysis."*

As noted elsewhere, there is an interdependence between methods, research questions and resources; fixed methods have their place, but can rarely be applied cleanly to address a real research problem (Furniss et al., 2011a), and may sometimes be used as labels to describe an approach that could not, in practice, conform exactly to the specified procedure.

As well as expertise in qualitative methods, the level of expertise in the study context can have a huge influence over the quality and kind of study conducted. When the study focuses on a widely used technology, or an activity that most people engage in, such as time management (e.g. Kamsin et al., 2012) or in-car navigation (e.g. Curzon et al., 2002), any disparity in expertise between researcher and participants is unlikely to be critical, although the researcher should reflect on how their expertise might influence their data gathering or analysis.

Where the study is of a highly specialised device, or in a specialist context, the expertise of the researcher(s) can have a significant effect on both the conduct and the outcomes of a study. At times, naivety can be an asset, allowing one to ask important questions that would be overlooked by someone with more domain expertise. At other times, naivety can result in the researcher failing to note or interpret important features of the study context.

Pennathur et al. (Pennathur et al. 2013, p.216) discuss this in the context of a study of technology use in an operating theatre:

> *"There was a possibility for bias and/or inconsistencies during identification of hazards in the [operating theatre] due to the involvement of observers with different expertise, and consequently the aspects that they may prioritise during observations."*

Domain expertise may also cause the researcher to become drawn into the on-going activity, potentially limiting their ability to record observations systematically – effectively becoming a practitioner rather than a researcher, insofar as these roles may conflict.

In preparing to conduct a study, it is important to consider the effects of expertise and to determine whether or not specific training in the technology or work being studied is required.

### 52.4.4 Other resources

There will be other resources and constraints that create and limit possibilities for the research design. These include the availability of equipment, funding – for example, for travel and to pay participants –, time, and suitable places to conduct research. Here, we briefly discuss some of these issues, while avoiding stating the obvious – variants on the theme of "don't plan to use resources that you don't have or can't acquire!".

Where a study takes place can shape that study significantly. Studies that take place within the context of work, home or other natural setting are sometimes referred to as 'situated' or 'in the wild' (e.g. Rogers, 2012). Studies that take place outside the context of work include laboratory studies – involving, for example, think-aloud protocol – and some interview studies – those that take place in 'neutral' spaces.

There are also intermediate points, such as the use of simulation labs, or the use of spaces that are 'like' the work setting, where participants have access to some, but not all, features of the natural work setting.

Observational studies most commonly take place 'in the wild', where the 'wild' may be a workplace, the home, or some other location where the activity of interest takes place, that is, the technology of interest is used. Interview studies may take place in the 'wild' or in another place that is comfortable for participants, quiet enough to record and ensure appropriate privacy, and safe for both participant and interviewer. Of course, there are also study types where researcher and participant are at-a-distance from each other, such as diary studies and remote interviews.

Rogers et al. (Rogers et al. 2011, p.227) discuss the uses of data recording tools including notes, audio recording, still camera, and video camera. All of these can be useful tools for data recording, depending on the situations in which data is being gathered. For instance, still photographs of equipment that has been appropriated by users, or a record of the locations in which technology was being used or how it was configured, provide a permanent record to support analysis, and to illustrate use in reports. As an example, Figure 52.2 shows how a home haemodialysis machine was marked up to remind the user to change a setting every time the machine was used.

Screen capture software can give a valuable record of user interactions with desktop systems. Particular qualitative methods such as the use of cultural probes

(Gaver and Dunne, 1999) or engaging participants in keeping video diaries, or testing ubiquitous computing technologies, may require particular specialist equipment for data gathering. When it comes to data analysis, coloured pencils, highlighter pens and paper are often the best tools for small studies. For larger studies, computer-based tools to support qualitative data analysis (e.g. NVivo or AtlasTI) can help with managing and keeping track of data, but require an investment of time to learn to use them effectively.

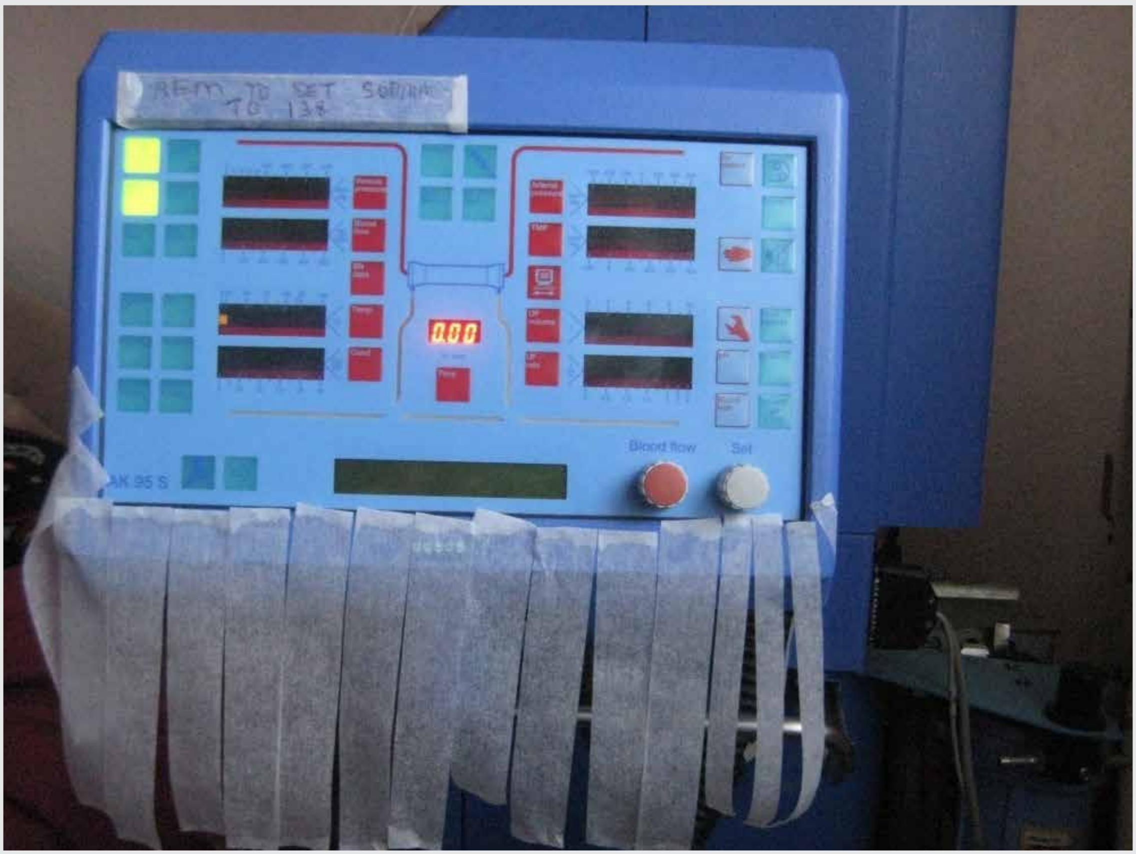





FIGURE 52.2: HHD machine with reminder "rem to set sodium to 136". The strips hanging at the bottom also show how the machine is being used as a temporary place to store cut strips for future use.

In addition to the costs of equipment, the other main costs for studies are typically the costs of travel and participant fees. Within HCI, there has been little discussion around the ethics and practicality of paying participant fees for studies. In disciplines where this has been studied, most notably medicine, there is little agreement on policy for paying participants (e.g. Grady et al., 2005; Fry et al., 2005). The ethical concerns in medicine are typically much greater than those in HCI, where the likelihood of harm is much lower. In HCI, it is common practice to recompense participants for their time and any costs they incur without making the payment, whether cash or gift certificates, so large that people are likely to participate just for the money.

## 52.5 ETHICS AND INFORMED CONSENT

Traditionally, ethics has been concerned with the avoidance of harm, and most established ethical clearance processes focus on this. 'VIP' is a useful mnemonic for the main concerns, **V**ulnerability, **I**nformed consent, and **P**rivacy:

- **V**ulnerability: particular care needs to be taken when recruiting participants from groups that might be regarded as vulnerable, such as children, the elderly, or people with a particular condition (illness, addiction, etc.).
- **I**nformed consent: where possible, participants should be informed of the purpose of the study, and of their right to withdraw at any time. It is common practice to provide a written information sheet outlining the purpose of the study, what is expected of participants, how their data will be stored and used, and how findings will be reported. Depending on the circumstances, it may be appropriate to gather either written or verbal informed consent; if written, the record should be kept securely, and separately from data. With the

growing use of social media, and of research methods making use of such data, from, for example, Twitter or online forums, there are situations where gathering informed consent is impractical or maybe even impossible; in such situations, it is important to weigh up the value of the research and means of ensuring that respect for confidentiality is maintained, bearing in mind that although such data has been made publicly available, the authors may not have considered all possible uses of the data and may feel a strong sense of ownership. If in doubt, discuss possible ethical concerns with local experts.

- **P**rivacy and confidentiality should be respected, in data gathering, management and reporting.

Willig (Willig 2008, p.16) lists informed consent, the avoidance of deception, the right to withdraw, debriefing, and confidentiality as primary considerations in the ethical conduct of research.

However, the work of the author and co-workers with clinicians and patients (Furniss et al., 2011b; Rajkomar & Blandford, 2012, Rajkomar et al., 2013) has highlighted the fact that ethics goes beyond these principles. It should be about doing good, not just avoiding doing harm. This might require a long-term perspective: understanding current design and user experiences to guide the design of future technologies. That long-term view may not directly address the desire of research participants to see immediate benefit.

What motivates an individual technology user to engage with research on the design and use of that technology? Corbin and Strauss (2008) suggest that one reason for participating in a study may be in order to make one’s voice heard.

This is not, however, universally the case. For example, in one of our studies of medical technologies (Rajkomar et al., 2013), participants were concerned to be seen as experts – because they might have had rights withdrawn if they were not – so it was *not* a benefit to them to have a chance to critique the design and usability of the system. For such participants, it may be about the 'common good': about being prepared to invest time and expertise for long-term benefits. For others, there is an indirect pay-back in terms of having their expertise and experience recognised and valued, or of being listened to, or having a chance to reflect on their condition or their use of technology.

If the study involves using a novel technology, there may well be elements of curiosity, opportunities to learn, and experiencing pleasure in people's motivations for taking part. Some people will be attracted by financial and similar incentives. There are probably many other complex motivations for participating in research. As researchers, we need to understand those motivations better, respect them, and work with them. Where possible, researchers need to 'repay' participants and others who facilitate research, and manage expectations where those expectations may be unrealistic – such as having a fully functioning new system within a few months.

Finally, Rogers et al. (Rogers et al. 2011, p.224) point out that the relationship between researcher and participants must remain "clear and professional". They suggest that requiring participants to sign an informed consent form helps in achieving this: true in some situations, but not in others, where verbal consent may be less costly and distracting for participants.

## 52.6 TECHNIQUES FOR DATA GATHERING

The most common techniques for data gathering in SSQSs are outlined below: observation; contextual inquiry; semi-structured interviews; think-aloud; focus groups; and diary studies. The increasing focus on the use of technologies while mobile, in the home, and in other locations are leading to yet more ways of gathering qualitative data. As Rode (Rode 2011, p.123) notes:

> *"As new technologies develop, they allow new possibilities for field-work – remote interviews, participant-observation through games, or blogs, or virtual worlds, and following the lives of one's informants via twitter."*

The possibilities are seemingly endless, and growing. The limit may be the imagination of the research team.

Whatever methods of data gathering are employed, it is wise to pilot test them before launching into extensive data gathering – both to check that the data gathering is as effective as possible and to ensure that the resulting data can be analysed as planned to address the purpose of the study. If the study design is highly iterative (e.g. using Grounded Theory as outlined in Section 52.1.5), then it is important to review the approach to data gathering before every data-gathering episode. If the data gathering and analysis are more independent, as in some other research designs, it is more important to include an explicit piloting stage to check that the approach to data gathering is working well: for example, to ensure that interview questions are effective or that participant instructions are clear).

### 52.6.1 Observation

There are many possible forms of observation, direct and indirect. Flick (Flick 2009, p.222) proposes five dimensions on which observational studies may vary:

- Covert vs. overt: to what extent are participants aware of being observed?
- Non-participant vs. participant: to what extent does the observer become part of the situation being observed?

- Systematic vs. unsystematic: how structured are the observation notes that are kept?
- Natural vs. controlled context: how realistic is the environment in which observation takes place?
- Self-observation vs. observation of others: how much attention is paid to the researcher's reflexive self-observation in data gathering?

In other words: there is no single right way to conduct an observational study. Indeed, the way it is conducted will often evolve over time, as the researcher's understanding of the context and ability to participate constructively and helpfully in it develop.

Flick (Flick 2009, p.223) identifies seven phases for planning an observational study:

- Selection of setting(s) for observation;
- Determining what is to be documented in each observation;
- Training of observers (see discussion on expertise, Section 52.4.3)
- Descriptive observations to gain an overview of the context;
- Focused observations on the aspects of the context that are of interest;
- Selective observations of central aspects of the context;
- Finish when theoretical saturation has been reached – i.e., when nothing further is being learned about the context.

These phases – particularly selective observations and theoretical saturation – convey a particular view of observation as developing a focused theory, much in the style of Grounded Theory (Section 52.1.5). Nevertheless, the broader idea of careful preparation for a study and recognition that the nature of observations will evolve over time are important for nearly all observational studies.

Willig (Willig 2008, p.28) discusses the nature of data gathering, including the importance of keeping detailed notes – such as near-verbatim quotations from participants and "concrete descriptions of the setting, people and events involved". She refers to these as "substantive notes", which may be supplemented by "methodological notes" – reflecting on the method applied in the research in practice – and "analytical notes", which constitute the beginning of data analysis (Section 52.7). She also notes that data collection and analysis may be more or less tightly integrated – a theme to which we return in Section 52.9.3.

### 52.6.2 Contextual Inquiry

Contextual inquiry (Beyer and Holtzblatt, 1998) is a widely reported method for conducting and recording observational studies in HCI, as a stage in a broader process of contextual design. According to Holtzblatt and Beyer (2013), "Contextual Design prescribes interviews that are not pure ethnographic observations, but involve the user in discussion and reflection on their own actions, intents, and values". In other words, contextual inquiry involves interleaving observation with focused, situated interview questions concerning the work at hand and the roles of technology in that work.

More importantly, Holtzblatt and Beyer (2013) present clear principles underpinning contextual design, and a process model for conducting design, including the contextual inquiry approach to data gathering. This includes a basic principle of the relationship between researcher and participants: that although the researcher may be more expert in human factors or system design, it is the participants who are experts in their work and in the use of systems to support that work.

Holtzblatt and Beyer (2013) present five models, flow, cultural, sequence, physical, and artefact, that are intermediate representations to describe work and the work context, and for which contextual inquiry is intended to provide data. Although contextual inquiry is often regarded as a component of contextual design, it has been applied independently as an approach to data gathering in research (e.g. Blandford and Wong, 2004).

### 52.6.3 Think Aloud

In contextual inquiry, the researcher is clearly present, shaping the data gathering through the questions he or she asks; in contrast, in a think-aloud study the researcher retreats into the background. Think aloud involves the users of a system articulating their thoughts as they work with a system. It typically focuses on the interaction with a particular interface, and so is well suited to identifying strengths and limitations of that interface, as well as the ways that people structure their tasks using the interface. Think aloud is most commonly used in laboratory studies, but also has a valuable role in some situated studies, as people demonstrate their use of particular systems in supporting their work (e.g. Makri et al., 2007).

Variants on the think-aloud approach are used in many disciplines, including cognitive psychology (e.g. Ericsson and Simon, 1980), education research (e.g. Charters, 2003) and HCI. Boren and Ramey (2000) conducted a review analysing the ways in which think aloud had been used in a variety of HCI studies, and conclude that, although most researchers cited Ericsson and Simon (1980) as their source for the method, the details of think alouds varied substantively from study to study.

Boren and Ramey (Boren and Ramey 2000, p.263) highlight four key principles from Ericsson and Simon (1980) to which a think aloud study should conform:

- Only 'hard' verbal data should be collected and analysed: "The only data considered must be what the participant attends to and in what order."
- Detailed instructions for how to think aloud should be given: "Encourage the participant to speak constantly 'as if alone in the room' without regard for coherency." They also recommend that

participants should have a chance to practise thinking aloud prior to the study.

- If participants fall silent, they should be reminded succinctly to verbalise their thoughts.
- Other interventions should be avoided, and attention should not be drawn to the researcher's presence. This is in stark contrast to the approach of contextual inquiry (Section 52.6.2).

Norgaard and Hornbaek (2006) conducted a study of how think aloud methods are used in practice, observing studies in seven different companies. They noted (p.271) that many of the studies did not conform to the guidelines above. For example, they included questions about people's perceptions, expectations, and interpretations during the TA study; exhibited a "tendency that evaluators end up focusing too much on already known problems"; and "evaluators seem to prioritize problems regarding usability over problems regarding utility".

In other words, as with most other data-gathering techniques, there are in practice many different ways to go about gathering data, and these are shaped by the interests of the researcher, the purpose of the study, and the practicalities of the situation.

One aspect of think aloud that has received little attention is how participants are instructed – not just in how to think aloud, but also in the tasks to be performed. Sometimes (e.g. Makri et al., 2007) these are naturalistic tasks chosen by the participants themselves, so that the researcher is essentially observing the participant completing a task that is part of, or aligned to, their on-going work. In other cases, tasks need to be defined for participants, and care needs to be taken to ensure that these tasks are appropriate, realistic and suitably engaging. While researcher-defined tasks are widely used in usability studies, they are less common in SSQSs, which are generally concerned with understanding technology use 'in the wild'.

## 52.6.4 Semi-structured Interviews

Think alouds are one way to gather verbal data from participants about the perceptions and use of technology; interviews are another widespread way of gathering verbal data. Interviews may be more or less structured: a completely structured interview is akin to a questionnaire, in that all questions are pre-determined, although a variety of answers may be expected; a completely un-structured interview is more like a conversation, albeit one with a particular focus and purpose. Semi-structured interviews fall between these poles, in that many questions – or at least themes – will be planned ahead of time, but lines of enquiry will be pursued within the interview, to follow up on interesting and unexpected avenues that emerge.

Interviews are best suited for understanding people's perceptions and experiences. As Flick (Flick 1998, p.222) puts it: "Practices are only accessible through observation; interviews and narratives merely make the accounts of practices accessible."

People's ability to self-report facts accurately is limited; for example, Blandford and Rugg (2002) asked participants to describe how they completed a routine task, and then to show us how they completed it. The practical demonstration revealed many steps and nuances that were absent from the verbal account: these details were taken for granted, so 'obvious' that participants did not even think to mention them.

Arthur and Nazroo (2003) emphasise the importance of careful preparation for interviews, and particularly the preparation of a "topic guide" (otherwise known as an interview schedule or interview guide). Their focus is on identifying topics to cover rather than particular questions to ask in the interview. It can be useful to have prepared important questions 'verbatim' – not because the question should then be asked rigidly as prepared, but because it identifies one way of asking it, which is particularly valuable if the interviewer has a ‘blank’ during the interview. Arthur and Nazroo advocate planning the topic guide within a frame comprising:

- Introduction;
- Opening questions;
- Core in-depth questions; and
- Closure.

This planning corresponds to the stages of an interview process as described by Legard et al. (2003), who present two views on in-depth interviewing. One starts from the premise that knowledge is ‘given’ and that the researcher’s task is to dig it out; although they do not use the term, this is in a positivist tradition. The other view is a constructivist one: that knowledge is created and negotiated through the conversation between interviewer and interviewee. Legard et al. (Legard et al. (2003) p.143) emphasize the importance of building a relationship, noting that the interviewer is a “research instrument”, but also that researchers need “a degree of humility, the ability to be recipients of the participant’s wisdom without needing to compete by demonstrating their own.”

They present the interview process as having six stages, all of which need to be planned for:

1. Arrival: the first meeting between interviewee and interviewer has a crucial effect on the success of the interview; it is important to put participants at their ease.
2. Introducing the research: this involves ensuring that the participant is aware of the purpose of the research, and has given informed consent, that they are happy to have the interview recorded, and understand their right to withdraw.

3. Beginning the interview: the early stages are usually about giving the participant confidence and gathering background facts to contextualize the rest of the interview.

4. During the interview: the body of the interview will be shaped by the themes of interest for the research. Participants are likely to be thinking in a focused way about topics that they do not normally consider in such depth in their everyday lives.

5. Ending the interview: Legard et al. emphasize the need to signal the end so that the participant can prepare for it and ensure there are no loose ends.

6. After the interview: participants should be thanked and told what will happen next with their data. Many participants think of additional things to say once the recorder is off, and these may be noted. Legard et al. emphasise the importance of participants being "left feeling 'well'" (Legard et al. (2003) p.146), as discussed in Section 52.5.

Legard et al. present various strategies for questioning, including the use of broad and narrow questions, avoiding leading questions, and making sure all questions are clear and succinct.

Within the core phase of interviewing, one technique to help with recall is the use of examples, asking people to focus on the details of specific incidents rather than generalizations. For example, the critical incident technique (Flanagan, 1954) can be used to elicit details of unusual and memorable past events, which in the context of HCI might include times when a technology failed or when particular demands were placed on a system.

A variant of this approach, the critical decision method (CDM), is presented in detail by Klein at al. (1989): in brief, their approach involves working with participants to reconstruct their thought processes while dealing with a problematic situation that involved working with partial knowledge and making difficult decisions. CDM helps to elicit aspects of expertise that are particularly well suited to studying technology use in high-pressure environments where the situation is changing rapidly and decisions need to be made, as in control rooms, operating theatres, and flight decks.

Charmaz (2006) describes an "intensive interview" as a "directed conversation". Her focus is on interviewing within grounded theory (Section 52.1.5), and on eliciting participants' experiences. She emphasizes the importance of listening, of being sensitive, of encouraging participants to talk, of asking open-ended questions, and not being judgemental. Although the participant should do most of the talking, the interviewer will shape the dialogue, steering the discussion towards areas of research interest while attending less to areas that are out of scope.

Charmaz emphasizes the "contextual and negotiated" (p.27) qualities of an interview: the interviewer is a participant in the shaping of the conversation, and therefore, the interviewer's role needs to be reflected in the outcome of a study. This is a theme to which we return in Section 52.9.2.

### 52.6.5 Focus groups

Focus groups may be an alternative to interviews, but have important differences. The researcher typically takes a role as facilitator and shaper, but the main interactions are between participants, whose responses build on and react to each other's. The composition of a focus group can have a great effect on the dynamic and outcome in terms of data gathered. Sometimes a decision will be made to gather data through focus groups to exploit the positive aspects of group dynamics; at other times, the decision will be more pragmatic.

For example, Adams et al. (2005) gathered data from individual practising doctors through interviews, partly because doctors typically had their own offices (a location for an interview), but also because they had very busy diaries. Each interview, therefore, had to be scheduled for a time when the participant was available (and many had to be delayed or rescheduled due to the demands of work). However, they gathered data from trainee nurses through focus groups because the nurses formed a cohort who knew each other reasonably well and often had breaks at the same time, so it was both easier and more productive to conduct focus groups than interviews.

### 52.6.6 Diary studies

Diary studies enable participants to record data in their own time – such as at particular times of day, or when a particular trigger occurs. Diary entries may be more or less structured; for example, the Experience Sampling Method (Consolvo and Walker, 2003) typically requires participants to report their current status in a short, structured form, often on a PDA / smart phone, whereas video diaries may allow participants to audio-record their thoughts, with accompanying video, with minimal structure.

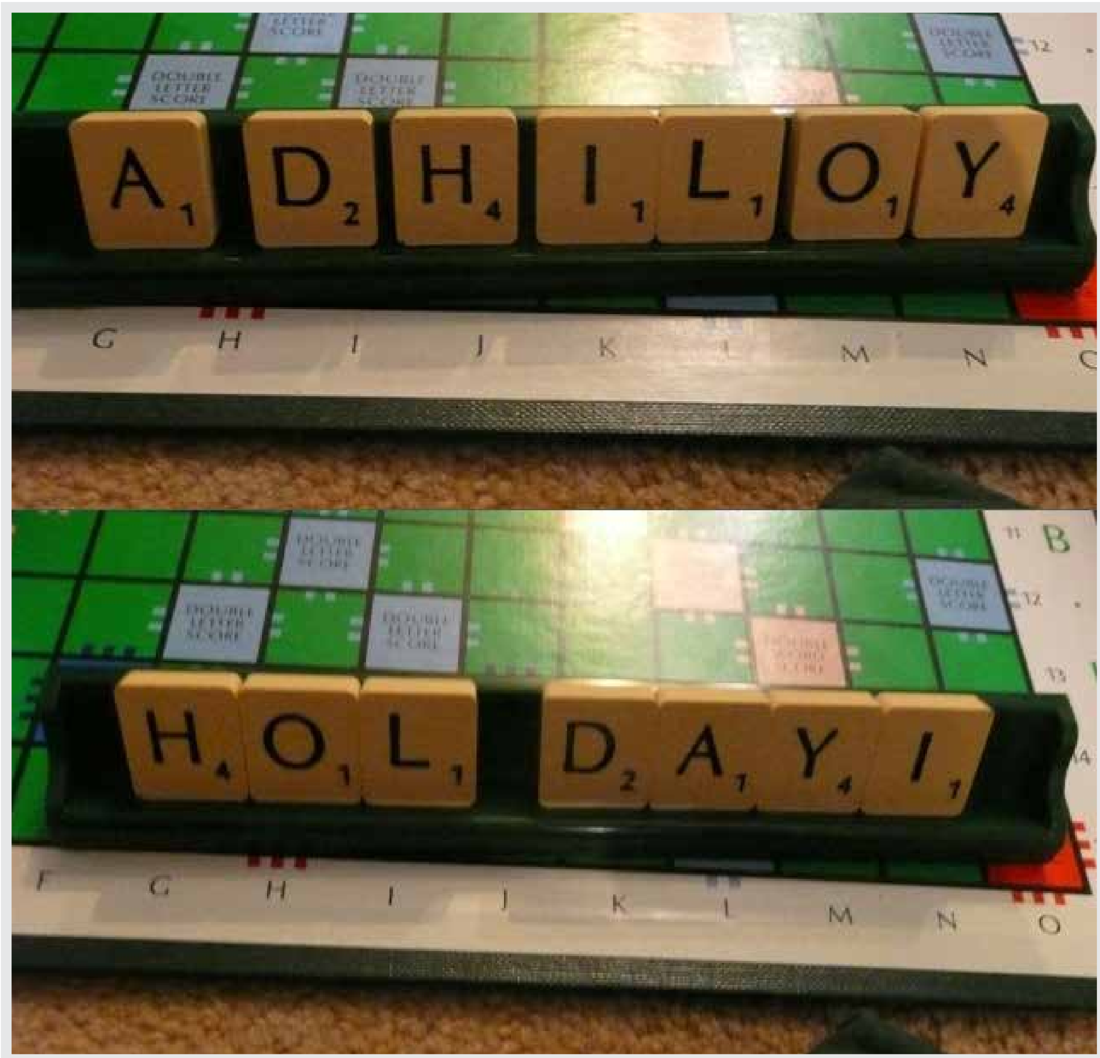
Kamsin et al. (2012) investigated people's time management strategies and tools using both interviews and video diaries. While interviews gave good insights into people's overall strategies and priorities, the immediacy of video diaries delivered a greater sense of the challenges that people faced in juggling the demands on their time and of the central role that email plays in many academics' time management.

### 52.6.7 Summary

Analysis can only work with the data that is collected. Therefore, it is important to gather the best possible data, working within the resources of the project (as discussed in Section 52.4). In some situations, data gathering and analysis are treated as being semi-independent from each other, with all analysis following the end of data gathering. In other situations, the two are interleaved – whether in the rich way advocated in GT, or by interleaving stages of data gathering and analysis as the study proceeds (e.g. as the theoretical focus develops, or as different data gathering methods are applied to address the problem from different angles – see discussion of triangulation in Section 52.10.2).

## 52.7 ANALYSIS

Most data for SSQSs exists in the form of field notes, audio files, photographs and videos. The first step of analysis is generally to transform these into a form that is easier to work with – such as transcribing audio, annotating or coding video. This may be done at different levels of detail; for example, selectively transcribing text that is directly relevant to the theme of the study through to a full transcription of all words, phatic utterances, pauses and intonations. The decision about which details to include should be guided by the purpose of the study, and hence the style of analysis to be completed. Some researchers choose to transcribe data themselves, as the very act of transcribing, and maybe making notes at the same time, is a useful step in becoming familiar with the data and getting immersed in it. Others prefer to pay a good typist to transcribe data, because, for example, they consider this a poor use of their time.




*Courtesy of Ann Blandford. Copyright: CC-Att-ND-3 (Creative Commons Attribution-NoDerivs 3.0 Unported).*

Figure 52.3: Reorganising Scrabble letters.

Similarly, people make different decisions about which tools to use for analysis. Decisions may be based on prior experience – such as having used Qualitative Data Analysis (QDA) tools such as ATLASti or NVivo, on the size and manageability of the dataset, and on personal preference. Any tool creates mediating representations between the analyst and the data, allowing the analyst to 'see' the data in new ways, just as reorganising Scrabble letters can help the player to 'see' words they may not have noticed previously (Figure 52.3). One researcher may choose to use a set of tables in a word processor to organise and make sense of data; another might create an affinity diagram (see Figure 52.4 for an example). Corbin and Strauss (Corbin and Strauss 2008, p.xi) emphasise that tools should "support and not 'take over' or 'direct' the research process", but note the value of tools in making analysis more systematic, contributing to reliability and an audit trail through the analysis.



Figure 52.4: Example of affinity diagramming with post-it notes, reproduced with permission from Stawarz (2012).

### 52.7.1 Different approaches to coding and iteratively analysing data

As noted above, an identifying feature of SSQSs is that they involve some form of coding of the data – i.e. of creating useful descriptors of units of data, such as single words, phrases, extended utterances, objects featuring in photographs, actions noted in videos, etc., and then of comparing and contrasting coded units to construct an analytical narrative based on the data. Grounded theory and thematic analysis (as outlined in Section 52.15 and Section 52.7.2) exemplify ways of coding data for analysis. There is a 'space' of approaches to coding qualitative data.

At one extreme, codes are simply 'buckets' in which to organise concepts identified in the data. Taking a utilitarian HCI approach, as a form of requirements gathering and system evaluation, CASSM (Blandford et al., 2008b; Blandford, in press) is a systematic approach to identifying the concepts that users are invoking when working with a system; where possible, these should be implemented in the system design (Johnson and Henderson, 2011).

A CASSM analysis involves gathering verbal data and classifying it in terms of user concepts. For example, in a study of ambulance control (Blandford et al., 2002), controllers were found to be working with two concepts both of which they referred to as 'calls': emergency calls being received; and the incidents to which those calls referred. The call management system they were working with at the time of the study allowed them to process emergency calls reasonably easily, but did not support incident management, which is important, particularly when a major incident occurs and many people call to report the same incident. A CASSM analysis is an SSQS, but the data analysis is simple, being mainly concerned with coding concepts by assigning them to 'buckets'.

Based in a positivist social sciences tradition, Miles and Huberman (Miles and Huberman 1994, p.64) advocate creating a preliminary list of codes prior to conducting fieldwork, and then refining it through analysis. They also advocate having multiple coders who can check that there is a shared understanding of codes to achieve "an unequivocal, common vision of what the codes mean and which blocks of data best fit which code."

At another point on the spectrum of approaches, Corbin and Strauss (Corbin and Strauss 2008, pp. 10 & 32) emphasise the centrality of the individual researcher in creating, interpreting and reporting the study:

> *"Concepts and theories are constructed by researchers out of stories that are constructed by research participants who are trying to explain and make sense out of their experiences and/or lives. ........... Sensitivity stands in contrast to objectivity. It requires that a researcher put him- or herself into the research. Sensitivity means having insight, being tuned in to, being able to pick up on relevant issues, events, and happenings in data. It means being able to present the view of participants and taking the role of the other through immersion in data."*

There are many available resources, such as those from the social sciences, that describe approaches to qualitative data analysis in detail. For example, Grbich (2013) presents over a dozen different approaches including: what she terms "classical ethnography", which is much more structured than the approach described in Section 52.1.4; three variants of grounded theory; cyber ethnography, focusing on internet use; and various approaches for analysing existing data. The challenge for the HCI researcher is to navigate their way through the space of possibilities, understanding the theoretical perspectives from which different authors are writing and constructing their own approach that is appropriate to the research question at hand, their own biases and competencies, and the resources available.

As well as the various approaches to GT, thematic analysis can be a valuable approach to analysing qualitative data, and exemplifies more of the space of possible approaches to analysis.

### 52.7.2 Thematic Analysis

In contrast to GT, where data collection and analysis are interleaved, thematic analysis assumes that a dataset already exists, and focuses attention on how that data might be analysed. Braun and Clarke (2006) argue that “thematising meanings” is a generic skill across qualitative methods and that thematic analysis builds directly on this skill. They contrast thematic analysis with qualitative techniques such as conversation analysis or interpretative phenomenological analysis, which are founded on a particular theoretical position and are typically applied in relatively tightly defined ways, but are rarely used in HCI. Rather, Braun and Clarke place thematic analysis in a ‘camp’ of techniques that can be applied across a range of theoretical positions, and that tries to steer a path between ‘anything goes’ unstructured analysis and an approach that is overly constrained. They make the obvious but important point (Braun and Clarke (2006) p.80), “What is important is that the theoretical framework and methods match what the researcher wants to know, and that they acknowledge these decisions, and recognise them *as* decisions.”

Braun and Clarke identify six phases of thematic analysis (Braun and Clarke (2006) pp. 87-88):

1. Familiarising with the data: simply reading and re-reading the data, making notes of ideas that spring to mind.

2. Generating initial codes: coding the entire dataset systematically and collating data that is relevant to each code. They define codes as labels that “identify a feature of the data (semantic content or latent) that appears interesting to the analyst”.

3. Searching for themes: gathering codes (and related data) into candidate themes for further analysis.

4. Reviewing themes: checking whether the themes work with the data and creating a thematic "map" of the analysis.
5. Defining and naming themes: refining the themes and the overall narrative iteratively.
6. Producing the report: which will, in turn require a further level of reflection on the themes, the narrative and the examples used to illustrate themes.

These phases represent an approach to iteratively deepening engagement with the data through layers of analysis.

Consistent with their overall flexible approach, Braun and Clarke (Braun and Clarke (2006), pp.88-89) are not prescriptive about whether an analysis should be informed (or driven) by a particular theory, or whether it should be driven by the analyst's interpretation of the data:

"Coding will, to some extent, depend on whether the themes are more 'data-driven' or 'theory-driven' – in the former, the themes will depend on the data, but in the latter, you might approach the data with specific questions in mind that you wish to code around."

This is a theme to which we return in Section 52.9.1.

## 52.8 REPORTING

As with any writing, the reporting of an SSQS has to be appropriate to the audience. If the study has been commissioned to deliver findings rapidly as part of a commercial development process, the reporting should be appropriately succinct and focused, whereas if it is part of a PhD thesis or other large academic project the reporting is more likely to focus on novel contribution and relationship to theory and previous literature.

There are many texts dedicated to the topic of how to write – whether in terms of the practicalities of getting and staying motivated, or of structuring text, or of structuring argument and addressing the intended audience. Thimbleby (2008) encourages the reader to write early and often, to draft and redraft, not to expect the first version to be the best. He observes that we are implicitly taught to write just once by tight deadlines and that the practice of writing and rewriting that is essential to refining ideas and communicating them effectively is a difficult one to develop.

The same principle is advocated, equally strongly, by Wolcott (2009), who goes so far as to discuss "shitty first drafts" (p.51), to be followed by better second drafts and excellent third drafts. Writing is much improved by getting feedback from others, so it is helpful to get into the practice of getting feedback even for early drafts. In some cases, it can be particularly helpful to get feedback – of a draft that is not too 'shitty' – from study participants, as a form of validation (see Section 52.10.2).

Within some research traditions, there are recognised structures that are widely conformed to; for example, in the sciences, a standard format is: aims, background, method, results, discussion, conclusion – and this format is sometimes advocated in traditions such as design where the material does not fit so naturally into this shape.

For qualitative studies, Wolcott (2009) argues strongly that this is not an effective structure, because presentation of background material delays the presentation of the key substance of the study. He argues that only essential background material should be included as part of the introduction, and that other related work should be introduced as needed through the narrative.

There is no one correct approach to structuring, and it can certainly be very challenging to fit the reporting of SSQSs into the standard 'scientific' structure. Unlike most quantitative research, where the researcher's understanding of the problem is unlikely to change much during a study, unless the hypothesis is poorly founded or the method inadequately planned or executed, during an SSQS the researcher is likely to learn much about the problem, and to 'see' it in different ways as understanding matures (Furniss et al., 2011a).

To take a simple example: a researcher doing a situated study in an unfamiliar environment is learning about the study context – beyond what can be read in published material about it – while doing the study, and yet the details of the context are part of the background to the research, and not, usually, research findings. The boundaries between method of analysis and results, between results and discussion, and between discussion and conclusions can seem just as blurred, particularly as understanding deepens through iterations of analysis.

If the final understanding and all the literature that relates to that understanding is presented up-front, the actual findings can seem underwhelming, even though they were not anticipated at the beginning. In such cases, it is often valuable to take the reader through highlights of the journey that the researcher has travelled so that the reader is exposed to some of the delight of discovery that the research team experienced – assuming that the researchers started from a sensible place.

For example, one study of the author's started with the purpose of understanding how underground train controllers use technology and work together, with the intention of conducting Distributed Cognition analyses of different control rooms to understand variability in design and practices. As data gathering proceeded, it became clear that commonalities were much greater than contrasts, and that a more interesting question was how the culture and use of technology has evolved to maintain safety. We, the researchers, decided to focus the background section of the report (Smith et al., 2009) on principles of train control, based on both literature and our early data gathering, and then to contextualise our findings in terms of the literature on resilience (such as Rochlin, 1999).

Understanding can develop, both as further data is gathered (e.g. Charmaz, 2006) and as new theoretical perspectives are encountered as ways of making sense of the data (e.g. Furniss et al., 2011a). Braun and Clarke (Braun and Clarke 2006, p.80) make an important point in noting that an "account of themes 'emerging' or being 'discovered' is a passive account of the process of analysis, and it denies the *active* role the researcher always plays in identifying patterns/themes".

Again, this highlights the fact that there are alternative ways of reporting, depending on the role(s) that the researchers perceive themselves as having played in the research process. Bringing the researcher into the narrative makes explicit their role, which may make the research findings seem less objective or authoritative than a more 'distanced' account.

Within HCI, the highly personalised account is rare, as it can seem to be at odds with the expectation that one is delivering an account that is appropriately objective to inform design. And yet there may be times, as, for example, when delivering rich accounts of user experience to help designers 'put themselves in the users' shoes', when such a personalised account is both more honest and more effective than a depersonalised one.

It is possible for accounts to be 'too honest', if that results in participants being disadvantaged through their participation in the study in any way. Lipson (1997) highlights pitfalls of reporting that apply generally to qualitative studies. In the context of HCI studies, the issues may be more focused around whether any non-participant readers would be able to identify any participants by reading the account, and how participants might feel about the way their contribution has been reported. For example, if a study includes a focus on errors that people make with technology then it needs to be reported in a way that does not make participants feel either stupid or vulnerable.

It is also possible to be too honest in reporting the journey at such a fine-grained level of detail that the reader is bored and cannot discern important

**information** from trivial details. It is important to be accountable while presenting the study at an appropriate level of abstraction.

In summary, there is no one correct way of reporting a qualitative study, in terms of ‘voice’ (e.g. to what extent the researcher is present in the narrative) or structure. The researcher should understand what is possible and what disciplinary rules they are violating if they choose to write in an unconventional way. As Wolcott (Wolcott 2009, p.66) puts it, “Before you begin to rock the boat, make sure you are in it”. What matters is that:

- the purpose of the account is clear, and that the account focuses on the purpose;
- essential information is presented, such as what was actually done (rather than delivering textbook accounts of methods), while respecting participants and their confidentiality;
- it addresses the intended audience (whether this be practitioners, other HCI researchers or specialists in the domain of the study);
- it is related well to relevant prior work, so that it is clear what is novel about this study;
- the findings are presented at a level of abstraction such that the novel contribution and the extent to which the findings generalize to other settings are clear; and
- it is coherent as a narrative.

As noted above, it is almost impossible to get writing right first time, and an iterative process of drafting, getting feedback from others, re-reading the draft critically (preferably after a break, to gain some distance from it), and re-drafting is essential. It is also important to know when to stop, though, because perfection is unachievable!

It is also worth considering whether there are multiple audiences or angles from the same study, which may be written up separately (while avoiding self-plagiarism by making sure that multiple reports address different questions within the overall study purpose). An informal test of self-plagiarism is whether each paper can cite the other and be clearly different.

Reporting multiple angles separately can be particularly advantageous when each paper needs to be fitted within a tight word or page limit. Tight constraints can, in practice, be very helpful for communicating effectively as it forces the author to think about what really matters in the narrative, to omit spurious information, and to write succinctly. However, writing well takes time: Pascal is widely credited with the apology, "I would have written a shorter letter, but I did not have the time."

## 52.9 FACTORS THAT SHAPE A STUDY

As well as resources, constraints and ethical considerations, there are various less tangible factors that also shape any study. These include the way that pre-existing theory can be used to inform data gathering, analysis, and/or reporting of a study, and also the biases, understanding and experience of the researcher(s) involved in the project. These and other factors together create a web of interdependencies.

### 52.9.1 The role of pre-existing theory in data gathering, analysis and reporting

No researcher is a *tabula rasa*: each comes to a study with pre-existing understanding, experience, interests, etc.Hertzum et al. (2001) consider this to be "chilling": that there is no objective, shared understanding, even with an activity as superficially simple as identifying usability difficulties from think-aloud data. If this is true for analysing pre-determined data with a pre-defined question, it clearly has an even greater effect on the research that is conducted if the researcher is shaping the entire study. For the individual, it may be difficult to identify or articulate many of the individual factors that shape the research they conduct, but one obvious factor is the role of theory in an SSQS. Theory may be most prominent towards the end of research project, may come into play during the analysis, or may shape the research from the outset.

Morse (1997) argues that the role of qualitative research is to deliver theory, whereas the role of quantitative research is to test theory. This is consistent with the focus of GT, in which theory is 'grounded' in data, and of thematic analysis, in which theory 'emerges', or is discovered / created, from data through analysis. Furniss et al. (2011a) present an account of a study of human factors (HF) practitioners' practices that includes an example of this relationship with theory. Based on the findings from a series of semi-structured interviews, a high-level theory was developed around the idea of 'downstream utility', which was seeded by the work of Wixon (2003) and developed into the use of a flowing river metaphor for describing how context-shaping factors influence the flow of a HF project.

Other researchers may seek a theoretical framework to help them make sense of data that seems very interesting but specific to the context of study. For example, Furniss et al. (2011a) were already familiar with the theory of distributed cognition (DC: Hollan et al., 2000 ; Furniss & Blandford, 2006 ). Although DC was not used for structuring data gathering, we thought it would be a useful framework for thinking about information flows in a HF project, providing a 'theoretical lens' on the analysis.

In contrast, Adams et al (2005) had to explicitly search for a theory to account for their findings. We had studied several different digital library (DL) deployment projects and found that making DLs more accessible to healthcare practitioners, by making them available through shared computers in the workplace, reduced their use when it was expected to increase it. Conversely, a project that had placed clinical librarians as members of multi-disciplinary care teams had increased use of DLs.

We explored theories such as DC and activity theory (Kaptelinin, 2013 ), but these did not help in accounting for our data. After some searching, we came across the theory of communities of practice (Wenger, 1998 ), which resonated with our data. This theory helped us to make sense of the data in a way that moved us from some interesting but idiosyncratic findings that were only relevant to our particular study contexts to findings that had some generalizability, and hence could be applied in other settings where new technology was being deployed.

Finally, others may intentionally structure data gathering and analysis around a particular theoretical framework, or are so steeped in a particular theoretical perspective that their approach to both data gathering and analysis is shaped by that perspective. For example, when studying people's strategies for information seeking, Makri et al. (2008) shaped their approach to data gathering and analysis around the work of Ellis et al. (1993, 1997) . Where this is done, it is important not to trust an existing theoretical framework unquestioningly, but to test and extend that framework: are there counter-examples that challenge the accuracy of the existing framework? Are there examples that go beyond the framework and introduce important extensions to it?

In summary: theories, of different kinds, can serve useful roles in SSQSs: in structuring the gathering and/or analysis of data and reporting of findings. A theory can be a 'lens', providing 'sensitizing concepts' that impose a partial structure on data that is gathered, helping to shape and focus data gathering. Similarly, a theory can help in shaping analysis, suggesting initial codes for analysing the data. In both of these cases, it is important not simply to accept a theory, but to test it, looking for evidence that might extend or contradict the established theory, while being mindful that there has to be a balance between power and generality in any theory. Make a theory too general and it typically loses analytical power, so not every extension to a theory is valuable.

### 52.9.2 The role of the researcher

A consideration that crops up repeatedly but is rarely discussed explicitly in reporting SSQSs is the role of the researcher, and the degree to which the researcher shapes the data gathering as it happens.

As noted in Section 52.4.3, the training and expertise of the researcher can have a significant influence on their role, particularly in terms of what data they are most sensitised to, and hence alert and responsive to. This will include their training in the research methods, in the principles and practices of using the technology of interest, in particular theories, and in the practices of participants, such as details of particular job roles).

Some kinds of studies, such as diary studies and think-aloud studies, typically involve relatively little engagement between researcher and participant, so that once the study is initiated the researcher is reliant on having designed it well and on participants recording data as anticipated, which is why good pilot testing is advisable!

Such studies are relatively easy to describe in terms of how participants were instructed. Interventions by the researcher may be planned, and may 'nudge' data gathering, but on the whole the approach does not evolve significantly during the study, and the role of the researcher is limited – it is, for example, likely that substituting one researcher for another would have little effect on what data is gathered. Similarly, an analysis of documents or online forum data, for instance, allow the researcher to be objective, at least in regard to the data gathering since all they are doing is selecting the data of focus.

The same is true of some interview and observational studies, particularly those that are relatively structured. It is important to be aware of, and reflect on, the effect that observation may have had on participants' behaviour. Heisenberg showed that even at a subatomic level the act of observing changed the state of the thing observed; it is not possible to plan the perfect study of a situation without both influencing and being influenced by that situation. The very act of observing, therefore, may influence participants' behaviour.

In interviewing, the role of the researcher is likely to be smallest when the interview is structured. Some observational studies involve the researcher acting as a 'fly on the wall', trying to minimise the effect of their presence on the activity being observed. It may be a reasonable approximation to assume that the presence of the researcher has little influence on the data that is gathered, although it is important to reflect on the likelihood that observational factors such as the Hawthorne effect, in which participants were found to perform better when being observed (Roethlisberger & Dickson, 1939 ), might have an impact on findings.

For such studies, it is simplest and clearest to take an objective view. If, for example the purpose of the study is to understand how a particular group of professionals use technology to support their working practices, and the implications therefore for design, then the data gathered might emphasise some features of the work and downplay others. It is unlikely, however, that the presence or behaviour of the researcher has a major influence on how participants behave or what they report.

For example, in our study (Section 52.4.2 & Section 52.7.1) of how ambulance controllers use technology to maintain awareness of the situation, both within the control room and in the outside world of ambulances and incidents (Blandford and Wong, 2004 ), it seemed reasonable to assume that the way we related to study participants had little influence on their performance as professionals. The way we collected, analysed and reported data therefore downplayed our role as researchers in that process. The focus was on being systematic and thorough in data gathering and analysis, and transparent in reporting, so that the reader could trace how conclusions related to data.

Such a view places the researcher outside the study setting as an objective observer of phenomena, assuming that those phenomena have a truth that is independent of the role of the researcher in gathering data.

More active participation brings the researcher into the frame, and increases their influence on the data being gathered. This is most obvious in studies involving action research, in which the researcher is intentionally intervening to assess the effects of interventions on perceptions, processes and outcomes (Kock, 2013 ). It is also likely to be the case where the researcher acts as a participant observer, playing an active role within the study context, and to a lesser extent in approaches such as Contextual Inquiry (Holtzblatt & Beyer, 2013 ), which bring the researcher into the observation / interview space, though data gathering is still shaped mainly by the activities being performed.

In other studies, the researchers and their relationship with participants is central to the research process: the relationship has a strong influence over what information is shared and how it is shared by participants, how it is interpreted by the researcher, and how it is reported and may be interpreted by the reader. For example, Rode et al. (2004) discuss their approach of exploring families' use of programmable technologies in the home by using fuzzy felt props as being "provocative", aiming to establish "rich dialog" with participants.

Semi-structured interviews inevitably bring in the interests of the researcher as well as the participant. To pretend that they are objective is to downplay the individuality of each researcher and of the relationship between researcher and participant. Willig (2008) emphasises the role of the interview as a dialogue between people. Where the interview strays into potentially sensitive areas, such as negative feelings around technology use, it is surely unethical to remain artificially detached from the setting. In such situations, it is impossible to substitute one researcher for another: the researcher is effectively a research instrument.

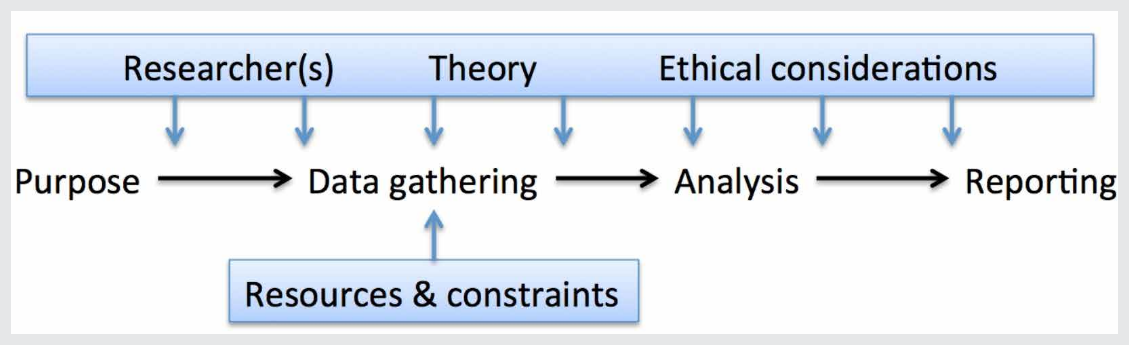
Where the research is truly exploratory, it is impossible to plan all the details of the study ahead of time and get them all right: the details have to evolve as understanding of the context and subject matter matures. This evolution is made explicit in the processes and ethos of Grounded Theory, but applies equally to other SSQSs that do not follow all the principles of GT. Such constructivist research demands reflexivity of the researcher in data gathering, analysis and reporting.

### 52.9.3 A web of considerations

As should already be apparent, there are many connections and interdependencies when designing, conducting and reporting SSQSs. And these phases of work are not generally distinct. Through engaging with the study setting, the researcher learns more about what is possible in terms of data gathering, and more about the nuances of the research question, so the purpose of the study may change, at least in subtle ways, as understanding evolves.

Unlike most quantitative studies, which can conveniently be treated as starting with a hypothesis and finishing with a conclusion – even if the truth is not quite that simple – many SSQSs are effectively journeys, in which the researcher travels alongside the participants, making discoveries that are shared through the reporting of the study. The focus for data gathering and analysis, therefore, may change, shaped by current understanding as the study proceeds. Furthermore, as discussed in earlier sections, the study is shaped by the individuals – researchers and participants – engaged in it, by any extant theory that is exploited at any stage in the study, by resources and constraints, and by ethical considerations.

Figure 52.5 shows an unrealistically simplified research process (which is the one often promoted by traditional reporting structures), highlighting the key stages of planning based on the purpose of the study: data gathering; analysis and reporting. All three are shaped by the expertise and understanding of the research team, any extant theory, ethical considerations, and resources and constraints, which typically have the greatest impact on data gathering.




*Courtesy of Ann Blandford. Copyright: CC-Att-ND-3 (Creative Commons Attribution-NoDerivs 3.0 Unported).*

FIGURE 52.5: An idealised process based on aims, methods, results, and discussion.

Figure 52.6 shows a process that is slightly closer to reality, with feedback and evolution in all stages as a study progresses, but still shaped by the same external factors. Data gathering and analysis may be more or less closely coupled. Early analysis may lead to revisions in the purpose of the study. The process of reporting often leads to new understanding of the problem. The overall purpose may be broken down into sub-questions that are best addressed through complementary studies involving different data gathering and analysis methods. These studies may be reported singly or together. Described in this way, the process can appear complicated and daunting, but in any particular instance the space of possibilities at any moment is not very great so, while every study is unique, it does not need to be unmanageable.

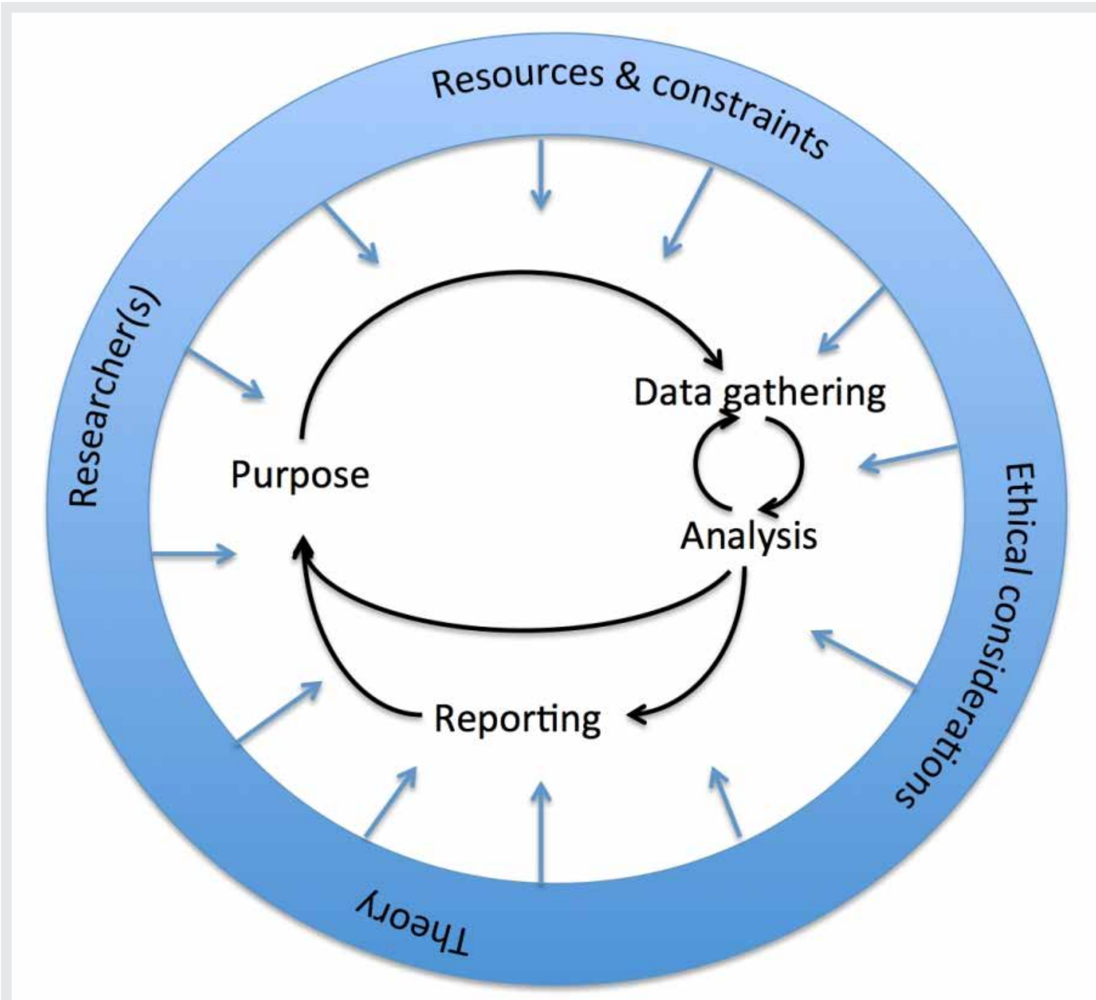





FIGURE 52.6: Closer to reality: a journey shaped by many factors.

One further factor that can have a great influence on findings is the participants, their motivations for taking part, and hence the data that is gathered in a study. The purpose of the study will determine who are ideal or possible participants, which may relate more-or-less directly to people's likely motivations for participating. This in turn should shape, and be shaped by, the recruitment strategy.

Participants will shape what data gathering and validation is possible and hence the quality of data analysis, which will determine the actual outcomes of the study. These outcomes should address the purpose of the study (Figure 52.7). As discussed above, all of these stages will also be constrained by ethical considerations.

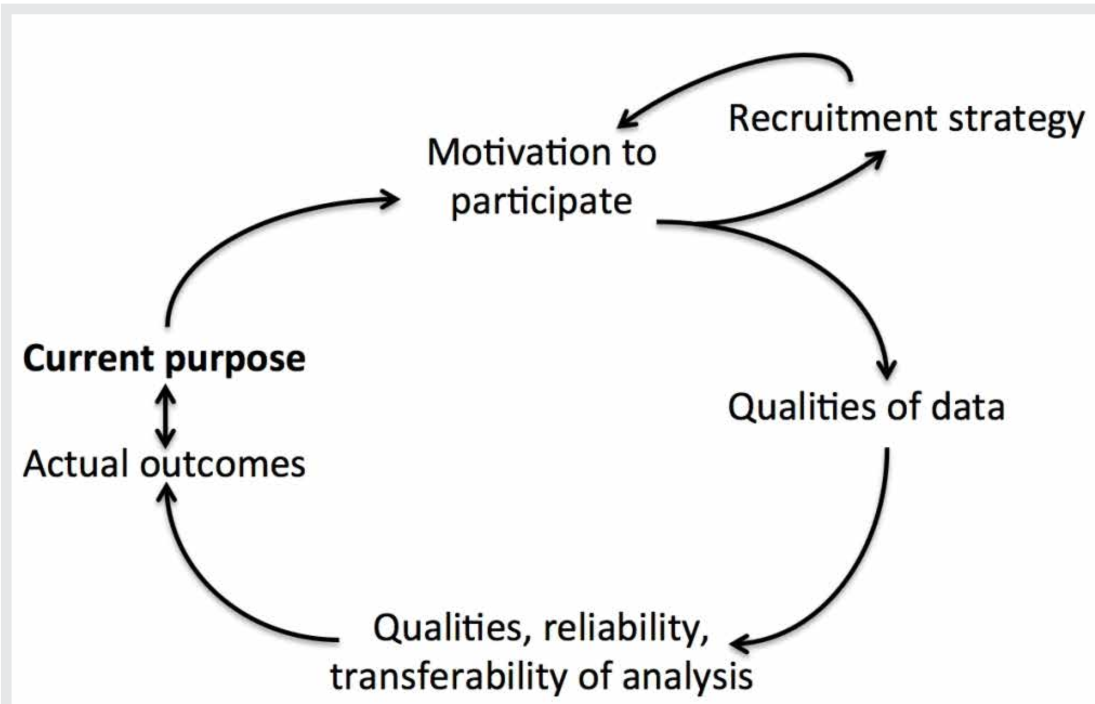





Figure 52.7: Interdependencies between the purpose of a study, recruitment of participants and outcomes – which should match the purpose.

## 52.10 ASSESSING AND ENSURING QUALITY IN QUALITATIVE RESEARCH

One of the challenges for qualitative researchers in HCI is that there is little consensus on what constitutes quality in qualitative research. Many reviewers adopt a particular stance, such as positivist or constructivist/interpretive, and immediately criticise research that does not conform to the expected paradigm. Arguably, on the one hand it is incumbent on the authors of a qualitative paper to present their approach and the rationale for it clearly, while on the other hand the reviewer has a responsibility to have appropriate expertise or an open mind, or to decline to review.

### 52.10.1 Quality criteria for constructivist research

In quantitative research, there are widely agreed criteria for quality such as internal validity, concerned with whether the experiment was properly conducted without confounding variables, and external validity, concerned with the generalizability of results. Kidder and Fine (Kidder and Fine 1987 , p.58) highlight some of the challenges in agreeing criteria in qualitative research by drawing an analogy between the work of a biographer and a qualitative researcher, quoting a psychohistorian: “When two quantitative researchers arrive at the same conclusions, we call it ‘reliability,’ but when two biographers write the same story we call it ‘plagiarism’.”

Implicitly, novelty and interest are assessment criteria in the latter case, though presumably other criteria, such as being justified on the basis of the available evidence, also apply. Within the space of SSQSs, there are different possible evaluative criteria.

Yardley (Yardley 2000 , p.219) proposes four essential characteristics of good qualitative research, which are also echoed by other authors:

- *Sensitivity to context*: e.g., taking account of previous relevant research, as well as ‘listening’ deeply to participants’ perspectives and being sensitive to ethical considerations. Klein and Myers (1999) emphasise the importance of enabling the reader to comprehend fully the context of the research.

- *Commitment and rigour*: e.g., engaging well with the topic and with participants, completing a thorough data collection, and conducting a thorough analysis.
- *Transparency and coherence*: e.g., making it clear how data was analysed and conclusions drawn. Similarly, Henwood and Pidgeon (1992) advocate keeping close to the data so that the link between data and conclusions is clear, and maintaining a 'paper trail' that is open to external audit to expose the layers of analysis.
- *Impact and importance*: e.g., articulating clearly both the theoretical and practical significance of findings. In HCI studies, this may, but not necessarily, include "implications for design" (Dourish, 2006 ). It may also include insight – that the study helps to understand work, interaction or experience with technology in a new way. Klein and Myers (1999) argue that importance is achieved through *abstraction and generalization* – i.e. relating the particulars of the study to general principles. Henwood and Pidgeon (1992) focus on *transferability*, arguing that researchers should report on the contextual aspects of the study that allow the reader to assess the sphere of relevance of findings. Transferability is similar, but not identical, to the idea of generalizability, being more focused on the question of how readily the findings from one study can be applied to a different context.

Focusing on constructivist research, Henwood and Pidgeon (Henwood and Pidgeon (1992) , p.105-108) list additional quality criteria, including:

- *Reflexivity*: the role of the researcher in the research should be recorded, and made apparent as appropriate. Klein and Myers (1999) advocate critical reflection on how data is "socially constructed" between researchers and participants.

- *Theoretical sampling and negative case analysis*: selecting cases that do not fit an emerging conceptual system helps to challenge assumptions. Henwood and Pidgeon consider this to be closely related to the constant comparative analysis advocated by Glaser, Strauss and others within the GT tradition. As well as being sensitive to contradictory evidence, Klein and Myers (1999) note the importance of being open to multiple interpretations (e.g. contradictory views of the same situation from different participants), yielding multiple narratives.

Coming from Information Systems research, Klein and Myers (1999) present principles based on a philosophical rationale. As well as echoing many of the criteria above, they also highlight the importance of:

- *The Hermeneutic Circle:* by this, they mean recognising that understanding is achieved by iterating between a focus on details and an understanding of the whole (similar to looking after the GOST of the study).
- *Suspicion:* i.e. being sensitive to possible systematic distortions in participants' narrative, for instance deriving from the way participants were recruited, or shaped by people's motivations for participating in the study.

These criteria for quality all depend on the researcher conducting the data gathering, analysis and reporting rigorously and honestly, and presenting the process with clarity and transparency.

### 52.10.2 External validation: inter-rater reliability, triangulation and respondent validation

There are also various approaches that give external validation of an analysis, which may be appropriate and feasible under some circumstances. These include employing multiple coders, triangulation of data sources, and respondent validation. These methods are typically built into the study design where they are used.

For some studies, the use of multiple coders, and maybe also measuring inter-rater reliability, is relevant. Miles and Huberman (Miles and Huberman 1994 , p.11) emphasise the importance of conclusions being verified, whether by reference back to field notes, achieving “intersubjective consensus” through discussion with colleagues, or replicating findings in another dataset. Miles and Huberman (1994) focus on the agreement of codes between multiple analysts – an approach that can be validated through measures of inter-rater reliability if coding is done independently.

Pennathur et al. (Pennathur et al. (2013) , p.207) work in a similar tradition, developing an approach to analysis that involves achieving group consensus for reconciling discrepancies between coders rather than computing inter-rater reliability. This requires that a set of codes has been previously agreed; in their case, these were based on the SEIPS model (Carayon et al., 2006 ) – i.e., on a particular theoretical perspective.

Having multiple independent coders of data and checking inter-rater reliability is appropriate for studies where codes and their meanings have been agreed and where the analysis and reporting relies heavily on those codes. It is not an appropriate way to validate a rich interpretive analysis.

In other situations, including many constructivist studies, it is possible to employ triangulation, which involves comparing multiple data sources or different methods of gathering data to corroborate findings. Mackay and Fayard (1997) argue that triangulation across scientific and design disciplines (introducing methods and theories from both) is particularly valuable in HCI. Rogers et al. (Rogers et al.2011 , p.225) list four different approaches to triangulation:

- *Triangulation of data*: data from different sources is compared; this helps with assessing the generalizability of findings.
- *Investigator triangulation*: different researchers collect and interpret the data; this is like the use of multiple coders as advocated by Miles and Huberman (1994) .
- *Triangulation of theories*: using different theoretical frameworks `as lenses on the data or findings.
- *Methodological triangulation*: employing different data gathering techniques can help to ensure that the outcome is not a simple function of the way that data was gathered.

Mays and Pope (2000) propose that, rather than supporting validation directly, triangulation encourages a more reflexive analysis of the available data. It depends on what form of triangulation is adopted as to how it can support data validation and give greater confidence in the findings, as outlined by Rogers et al. (2011) .

Another widely discussed and used approach is respondent validation, or 'member checking,' in which study participants are invited to review the study findings to validate the researchers' interpretation of the data. A variant on this is to have other representatives of the same group – i.e., people like the participants – review the findings. While some (e.g. Lincoln and Guba, 1985 ) regard this as a strong check, others (e.g. Mays and Pope, 2000 ) highlight weaknesses in the approach, including dealing with discrepancies in the responses of participants, which effectively represent new data to be analysed, and managing the different priorities and focuses of participants and researchers.

Rather than conducting standard respondent validation, Henwood and Pidgeon (1992) suggest that negotiating interpretations with participants may sometimes be an effective approach to validating interpretations. However, they also recognise that neither of these approaches is universally applicable – as when, for example, participants have reason to object to a particular interpretation of the data.

A further, informal, check is face validity: do the findings of the study make sense? Are they credible? On its own, face validity is a very weak test, and should always be viewed with a critical eye, but the converse can be helpful: findings that lack face validity are rightly viewed with suspicion.

Barbour (2001) suggests that in healthcare research there is a tendency towards what I would term a 'checklist mentality': what she calls "technical fixes" are being required by funders and/or reviewers to ensure the rigour of qualitative research. She highlights five such fixes: purposive sampling; grounded theory; multiple coding; triangulation; and respondent validation. For each, she discusses the potential benefits: reducing bias; supporting original theorising; enhancing interrater reliability; checking internal validity; and checking researchers' interpretations, respectively. However, she also highlights pragmatic limitations of each approach in practice, and argues (Barbour (2001) p.1117), "They can strengthen the rigour of qualitative research only if they are embedded in a broad understanding of qualitative research design and data analysis."

### 52.10.3 Building quality and value indicators into reporting

In HCI, the question of transferability, generalizability or scope of findings is important: if design decisions for future systems are to be based on findings from qualitative studies, there has to be confidence in their broader applicability, or an understanding of how broad their applicability is. Some confidence in generalizability can come from relating findings to established theory or by triangulating findings across different data sources (Section 52.10.2): if the findings from the current study are consistent with those from other studies – whether represented directly in their findings or through theory that abstracts from findings – then that is one source of confidence.

If findings differ in interesting ways from theory or previous studies then that merits further discussion: is this because of some important difference in study conditions such as taking place in a different kind of setting, or working with a different user population? Otherwise, how else might the discrepancy be accounted for? Alternatively, particularly where there is no relevant prior theory or data, the findings from a qualitative study might indicate the need for further research to test those findings.

The quality of studies varies for many reasons, often linked to what is possible with the available resources, including the experience and expertise of the researcher(s), the time available, or the ease of recruiting an appropriate group of participants. While it is obviously important to conduct the best possible study with the resources available, it is also important to report findings in a way that makes it possible for the reader to assess the quality of the research. The reader should be able to answer questions such as:

- What confidence do I have in the results and conclusions of this study? What is the evidence to support my judgement?
- What can I learn from this study (relative to what was known before)? What is novel?
- How can I build on this study? – whether on the methods, the findings, or gaps in knowledge that it has exposed.

For reviewers of papers reporting SSQSs, the question is perhaps more basic: is this paper worth publishing? It is probably impossible to conduct a 'perfect' qualitative study: with more resources, it is almost always possible to do a better job. So the question is: what's good enough?

In the preceding sections, I have outlined some of the dimensions on which qualitative studies vary. In this section I have reviewed some perspectives on quality in qualitative studies, emphasising that there is not a 'one size fits all' approach, but that methods and approaches to quality control and validation should be coherent, and appropriate to the purpose, resources and methods of the study.

## 52.11 A CHECKLIST FOR DESIGNING AND REPORTING SSQSS

Drawing all these themes together, we can identify a 'space' of considerations for designing, conducting and reporting SSQSs. The actual details of a study design involve nuanced interrelationships between design aspects, and only some possible combinations of design decisions are coherent. Established named methods, such as Grounded Theory or Thematic Analysis, occupy regions of this space, but there are other possible designs of SSQSs that address important research questions and work with the available resources and constraints to deliver important HCI findings, and that do not have a particular name.

Any study demands sensitivity and adaptation to the situation. It also needs to be coherent and clear: you need to 'look after your GOST'. Clearly presented methods abound in textbooks and papers. They are there to be adopted and adapted to be fit for purpose.

Table 1 presents a checklist of questions that should be considered in the design, conduct and reporting of SSQSs. As discussed above (Section 52.8), there will be decisions that need to be made but that are not an essential part of reporting: that should focus on key decisions, key changes in plan and their likely impact on findings.

| | Planning and conducting study | Additional considerations for reporting |
|---|---|---|
| | Planning and conducting study | Additional considerations for reporting |
| Purpose (§52.3) | What is the purpose of the study?<br>Why is it an important study to conduct?<br>What gap in knowledge is it filling? | Did the purpose of the study change?<br>If so, why and how?<br>What are the novel and important findings?<br>Why do they matter to the reader? |
| Resources and constraints (§52.4) | What resources do you have to work with?<br>What constraints limit possibilities? | Were there any novel features of the way resources were used (e.g. new technology probes or innovative use of social media)?<br>Did the availability of resources (e.g. time) limit what was possible in important ways? |

| | | |
|---|---|---|
| *Researcher attributes and role* (§52.4.3 and §52.9.2) | How many people are in the research team, and what are their roles?<br>What knowledge and expertise does each researcher bring to the study?<br>What training will each receive?<br>To what extent will the researcher participate in the situation being observed (for observational studies)?<br>What is the intended relationship between researcher(s) and participants? | Are there attributes of the research team that will have influenced the study in important ways?<br>What role(s) did the researcher(s) play in the study setting?<br>How did the relationship that was established with each participant influence the data that was gathered (if it's possible to tell)? |
| *Advocacy (§52.4.1)* | Do you need advocate(s) within the study setting? How will you identify and work with them? | Who did you work with, and what was their influence (e.g. in terms of helping to refine research questions or recruit participants)? |

| | | |
|---|---|---|
| *Participant recruitment (§52.4.2)* | What is the approach to sampling participants? How (practically) will participants be recruited?<br>What are inclusion and exclusion criteria, and how might they evolve over the course of the study?<br>What is the anticipated relationship between researcher(s) and participant? | How were participants recruited in practice?<br>Were there compromises that needed to be made, and what is the likely impact of this on the quality, reliability or generalizability of findings?<br>What roles did researcher(s) and participant(s) take in the study? |
| *Location and intervention (§52.4.4)* | Where will the study take place? What forms of intervention are planned (e.g. introduction of novel prototype designs)? How naturalistic is the study? | Did the location(s) in which the study took place, or any interventions, influence outcomes in any important ways? |
| *Role of theory (§52.9.1)* | To what extent, and how, will theory play a role in data gathering, analysis and/or reporting? | How, if at all, did established theory shape the study? How do the findings relate to established theory? |

| | | |
|---|---|---|
| Ethical considerations (§52.5) | Are there important ethical considerations that need to be addressed? How will you ensure that participants benefit as far as possible from participation?<br>What will participants be told about the study?<br>How will data be stored and anonymised? | Did ethical considerations shape the study in important ways? If so, how?<br>How were participants informed about the study and what would be done with the data? |
| Techniques for data gathering (§52.6) | How will data be gathered (interviews, observation, etc.)? How will it be recorded?<br>If multiple methods are to be used, how will they be sequenced and coordinated?<br>What data will be gathered? How structured will the data gathering be? Will it be informed by theory? Define a protocol for observation, or a semi-structured interview script, or participant instructions (for think-aloud).<br>How will data gathering be timed (e.g. to sample particular kinds of activity)? | How was data gathered in practice? How did data gathering change over the course of the study (if at all)?<br>How were participants instructed (e.g. for a think-aloud study)? |

| | | |
|---|---|---|
| *Interleaving recruitment, data gathering and analysis* | How interdependent will participant recruitment, data gathering and analysis be? | How was data gathering and analysis interleaved (if at all)? How did early analysis shape later data gathering? |
| Analysis of data (§52.7) | How will data be analysed?<br>At what level of detail will transcription take place? What tools will be used to support analysis? Are codes pre-determined or identified through analysis? Are they agreed by a team? If there are multiple coders, is their coding independent or negotiated? Will participants be involved in analysis and / or validation?<br>If the analysis is individual and reflexive, what steps will the researcher take to ensure the validity of findings? | How was data analysed in practice? How iterative and reflexive was the analysis process? How was data validated in practice? |
| Reporting (§52.8) | Who is the audience? How will findings be reported? | What is novel? What is important? What is the evidence to support the claims being made? |

TABLE 52.1: A checklist for planning and reporting SSQSs.

This checklist has doubtless overlooked some important decisions in planning and in reporting. Please do add any omissions as comments on this chapter.

## 52.12 CONCLUSION

Wolcott (Wolcott 2009 , p.36) quotes a biologist, Paul Weiss, as claiming, "Nobody who followed the scientific method ever discovered anything interesting." Whether or not that is strictly true, one of the delights of exploratory qualitative studies is that they frequently deliver interesting, even surprising, findings.

In healthcare, there is a persistent view that randomised controlled trials are the 'gold standard' that defines criteria for quality in research (e.g. Concato et al., 2000 ). Although HCI has been less blighted by such a hierarchical view of research designs, there has nevertheless been a tendency to dismiss some forms of qualitative research as lacking rigour. While this is true of some studies, at other times it seems to be due to limited understanding of the culture, principles and processes of qualitative research.

All research demands trust: that the researcher did what they claim to have done, with integrity, and that the presentation is as accurate as possible. Because SSQSs are suitable for addressing a range of research questions, and because every study setting is different, there is not a 'one size fits all' method: methods need to be adapted to work with the resources and constraints of the project. Named methods should not be used as 'bumper stickers'; it is important to describe what was actually done at an appropriate level of detail to enable others to judge the quality of a study, and the implications for future research and practice.

My aim in this chapter has been to lay out a space of possibilities and considerations for Semi-Structured Qualitative Studies in HCI, and to provide pointers to literature where further details can be found. Not every qualitative research project is an ethnography or a GT. Not every project results in implications for design. There are many possible research questions, study designs and study outcomes. The challenge is to ensure that studies are of high quality, and outcomes of interest and value.

## 52.13 ACKNOWLEDGMENTS

I am indebted to the many colleagues, post-doctoral researchers, PhD students and MSc students with whom I have engaged on this exploration of qualitative methods to date, too numerous to name. Particular thanks, however, to Anne Adams, Dominic Furniss, Aisling O'Kane, Sheila Pontis, Atish Rajkomar, Kathy Stawarz and Chris Vincent for feedback on a draft of this chapter. I take responsibility for any remaining weaknesses. The work reported here has been funded by EPSRC, ESRC and SSHRC, including on-going grants EP/G059063/1 and EP/H042741/1.

## 52.14 REFERENCES

*Adams, Anne, Blandford, Ann and Lunt, Peter (2005): Social Empowerment and Exclusion: A case study on Digital Libraries. In ACM Transactions on CHI, pp. 174-200*

*Adams, Anne, Lunt, Peter and Cairns, Paul (2008): A qualitative approach to HCI research. In: Cairns, Paul and Cox, Anna (eds.). "Research Methods for Human-Computer Interaction". Cambridge University Presspp. 138-157*

*Anderson, R. J. (1994): Representations and Requirements: The Value of Ethnography in System Design. In Human-Computer Interaction, 9 (2) pp. 151-182*

*Atkinson, Rowland and Flint, John (2001): Accessing hidden and hard-to-reach populations: Snowball research strategies. In Social research update, 33 (1) pp. 1-4*

*Attfield, Simon and Blandford, Ann (2011): Making Sense of Digital Footprints in Team-Based Legal Investigations: The Acquisition of Focus. In Human Computer Interaction, 26 (1) pp. 38-71*

*Attfield, Simon, Fegan, Sarah and Blandford, Ann (2008): Idea Generation and Material Consolidation: Tool Use and Intermediate Artefacts in Journalistic Writing. In Cognition, Technology & Work, 11 (3) pp. 227-239*

*Barbour, Rosaline S. (2001): Checklists for improving rigour in qualitative research: a case of the tail wagging the dog?. In BMJ: British Medical Journal, 322 (7294) pp. 1115-1117*

*Beyer, Hugh and Holtzblatt, Karen (1998): Contextual design: defining customer-centered systems. San Francisco, Elsevier*

*Beyer, Hugh and Holtzblatt, Karen (2013): Contextual Design. In: Soegaard, Mads and Dam, Rikke Friis (eds.). "The Encyclopedia of Human-Computer Interaction, 2nd Ed". Aarhus, Denmark: The Interaction Design Foundation. Available online at*

*Blandford, Ann (2013): Eliciting People's Conceptual Models of Activities and Systems. In International Journal of Conceptual Structures and Smart Applications, 1 (1)*

*Blandford, Ann and Rugg, Gordon (2002): A case study on integrating contextual information with analytical usability evaluation. In International Journal of Human-Computer Studies, 57 (1) pp. 75-99*

*Blandford, Ann and Wong, B. L. William (2004): Situation awareness in emergency medical dispatch. In International Journal of Human-Computer Studies, 61 (4) pp. 421-452*

*Blandford, Ann, Wong, William B. L., Connell, Iain and Green, Thomas (2002): Multiple Viewpoints On Computer Supported Team Work: A Case Study On Ambulance Dispatch. In: Faulkner, Xristine, Finlay, Janet and DÃ©tienne, FranÃ§oise (eds.) Proceedings of the HCI02 Conference on People and Computers XVI September 18-20, 2002, Pisa, Italy. pp. 139-156*

*Blandford, Ann, Adams, Anne, Attfield, Simon, Buchanan, George, Gow, Jeremy, Makri, Stephann, Rimmer, Jon and Warwick, Claire (2008a): The PRET A Rapporter framework: Evaluating digital libraries from the perspective of information work. In Information Processing & Management, 44 (1) pp. 4-21*

*Blandford, Ann, Green, T. R. G., Furniss, Dominic and Makri, Stephann (2008b): Evaluating system utility and conceptual fit using CASSM. In International Journal of Human-Computer Studies, 20 (6) pp. 393-409*

*Boren, T. and Ramey, J. (2000): Thinking aloud: reconciling theory and practice. In Professional Communication, IEEE Transactions on, 43 (3) pp. 261-278*

*Braun, Virginia and Clarke, Victoria (2006): Using thematic analysis in psychology. In Qualitative research in psychology, 3 (2) pp. 77-101*

*Buckley, Brian, Murphy, Andrew W., Byrne, Molly and Glynn, Liam (2007): Selection bias resulting from the requirement for prior consent in observational research: a community cohort of people with ischaemic heart disease. In Heart, 93 (9) pp. 1116-1120*

*Button, Graham and Sharrock, Wes (2009): Studies of Work and the Workplace in HCI: Concepts and Techniques. Morgan and Claypool Publishers*

*Cairns, Paul and Cox, Anna (2008): Research Methods for Human-Computer Interaction. Cambridge University Press*

*Carayon, Pascale, Karsh, Ben-Tzion, Brennan, Patricia Flatley, Gurses, Ayse P., Hundt, Ann Schoofs, Alvarado, Carla J. and Smith, Maureen (2006): Work system design for patient safety: the SEIPS model. In Quality and Safety in Health Care, 15 (1) pp. 150-158*

*Charmaz, Kathy (2008): Grounded Theory. In: Smith, Jonathan A. (ed.). "Qualitative psychology: a practical guide to research methods". Sage*

*Charmaz, Kathy (2006): Constructing Grounded Theory: A Practical Guide through Qualitative Analysis. Pine Forge Press*

*Charters, E. (2013). The Use of Think-aloud Methods in Qualitative Research An Introduction to Think-aloud Methods*

*Concato, John, Shah, Nirav and Horwitz, Ralph I. (2000): Randomized, Controlled Trials, Observational Studies, and the Hierarchy of Research Designs. In New England Journal of Medicine, 342 (25) pp. 1887-1892*

*Consolvo, Sunny and Walker, Miriam (2003): Using the Experience Sampling Method to Evaluate Ubicomp Applications. In IEEE Pervasive Computing, 2 (2) pp. 24-31*

*Corbin, Juliet and Strauss, Anselm (2008): Basics of qualitative research: Techniques and procedures for developing grounded theory. Sage*

*Crabtree, Andrew, Rodden, Tom, Tolmie, Peter and Button, Graham (2009): Ethnography considered harmful. In: Proceedings of ACM CHI 2009 Conference on Human Factors in Computing Systems 2009. pp. 879-888*

*Curzon, Paul, Blandford, Ann, Butterworth, Richard and Bhogal, Ravinder (2002): Interaction design issues for car navigation systems. In: Sharp, Helen, Chalk, Pete, LePeuple, Jenny and Rosbottom, John (eds.) Proceeding of HCI 2002 September 2-6, 2002, London, United Kingdom. pp. 38-41*

*Dourish, Paul (2006): Implications for design. In: Proceedings of ACM CHI 2006 Conference on Human Factors in Computing Systems 2006. pp. 541-550*

*Ellis, David, Cox, Deborah and Hall, Katherine (1993): A Comparison of the Information-seeking Patterns of Researchers in the Physical and Social Sciences. In Journal of Documentation, 49 (4) pp. 356-369*

*Ericsson, K. A. and Simon, Herbert A. (1984): Protocol Analysis: Verbal Reports as Data. Cambridge, MA, MIT Press*

*Flanagan, John C. (1954): The critical incident technique. In Psychological Bulletin, 51 (4) pp. 327-358*

*Flick, Uwe (2009): An introduction to qualitative research. Sage*

*Fry, Craig L., Ritter, Alison, Baldwin, Simon, Bowen, Kathryn, Gardiner, Paul, Holt, T., Jenkinson, Rebecca and Johnston, Jennifer (2005): Paying research participants: a study of current practices in Australia. In Journal of medical ethics, 31 (9) pp. 542-547*

*Furniss, Dominic and Blandford, Ann (2006): Understanding Emergency Medical Dispatch in terms of Distributed Cognition: a case study. In Ergonomics, 49 (12) pp. 1174-1203*

*Furniss, Dominic, Blandford, Ann and Mayer, Astrid (2011): Unremarkable errors: low-level disturbances in infusion pump use. In: Proceedings of the 25th BCS Conference on Human-Computer Interaction BCS-HCI 11 July 4-8, 2011, Newcastle, United Kingdom. pp. 197-204*

*Furniss, Dominic, Blandford, Ann and Curzon, Paul (2011): Confessions from a grounded theory PhD: experiences and lessons learnt. In: Proceedings of ACM CHI 2011 Conference on Human Factors in Computing Systems 2011. pp. 113-122*

*Gaver, William W. and Dunne, Anthony (1999): Projected Realities: Conceptual Design for Cultural Effect. In: Altom, Mark W. and Williams, Marian G. (eds.) Proceedings of the ACM CHI 99 Human Factors in Computing Systems Conference May 15-20, 1999, Pittsburgh, Pennsylvania. pp. 600-607*

*Giorgi, Amedeo and Giorgi, B. (2008): Phenomenology. In: Smith, Jonathan A. (ed.). "Qualitative psychology: a practical guide to research methods". Sagepp. 26-52*

*Glaser, Barney and Strauss, Anselm (2009): The discovery of grounded theory: Strategies for qualitative research. Transaction Books*

*Grady, Christine, Dickert, Neal, Jawetz, Tom, Gensler, Gary and Emanuel, Ezekiel (2005): An analysis of U.S. practices of paying research participants. In Contemporary clinical trials, 26 (3) p. 365–375*

*Grbich, Carol (2013): Qualitative data analysis: An introduction. Sage*

*Hartswood, Mark, Procter, Rob, Rouncefield, Mark and Slack, Roger (2003): Making a Case* *in Medical Work: Implications for the Electronic Medical Record. In Computer Supported Cooperative Work, 12 (3) pp. 241-266*

Heath, Christian and Luff, Paul (1991): *Collaborative Activity and Technological Design: Task Coordination in London Underground Control Rooms. In: Bannon, Liam, Robinson, Mike and Schmidt, Kjeld (eds.) ECSCW 91 - Proceedings of the Second European Conference on Computer-Supported Cooperative Work September 24-27, 1991, Amsterdam, Netherlands. pp. 65-80*

Henderson, Austin and Johnson, Jeff (2011): *Conceptual Models: Core to Good Design. Morgan and Claypool Publishers*

Henwood, Karen L. and Pidgeon, Nick F. (1992): *Qualitative research and psychological theorising. In British Journal of Psychology, 83 (1) pp. 97-111*

Hertzum, Morten and Jacobsen, Niels Ebbe (2001): *The Evaluator Effect: A Chilling Fact About Usability Evaluation Methods. In International Journal of Human-Computer Interaction, 13 (4) pp. 421-443*

Hollan, James D., Hutchins, Edwin and Kirsh, David (2000): *Distributed Cognition: Toward a New Foundation for Human-Computer Interaction Research. In ACM Transactions on Computer-Human Interaction, 7 (2) pp. 174-196*

Hughes, John, King, Val, Rodden, Tom and Andersen, Hans (1994): *Moving Out from the Control Room: Ethnography in System Design. In: Proceedings of the 1994 ACM conference on Computer supported cooperative work October 22 - 26, 1994, Chapel Hill, North Carolina, United States. pp. 429-439*

Hutchins, Edwin (1995): *Cognition in the wild. Cambridge, Mass, MIT Press*

Kamsin, Amirrudin, Blandford, Ann and Cox, Anna L. (2012): *Personal task management: my tools fall apart when I'm very busy!. In: CHI12 Extended Abstracts on Human Factors in Computing Systems May 5-10, 2012, Austin, USA. pp. 1369-1374*

Kaptelinin, Victor (2013): *Activity Theory. In: Soegaard, Mads and Dam, Rikke Friis (eds.). "The Encyclopedia of Human-Computer Interaction, 2nd Ed". Aarhus, Denmark: The Interaction Design Foundation. Available online at*

Kidder, Louise H. and Fine, Michelle (1987): *Qualitative and quantitative methods: When stories converge. In New directions for program evaluation, 1987 (35) pp. 57-75*

Kindberg, Tim, Spasojevic, Mirjana, Fleck, Rowanne and Sellen, Abigail (2005): *The ubiquitous camera: an in-depth study of camera phone use. In IEEE Pervasive Computing, 4 (2) pp. 42-50*

Klein, Heinz K. and Myers, Michael D. (1999): A set of principles for conducting and evaluating interpretive field studies in information systems. In MIS Quarterly, 23 (1) pp. 67-93

Klein, Gary A., Calderwood, Roberta and MacGregor, Donald (1989): Critical decision method for eliciting knowledge. IEEE Transactions on Systems. In Man and Cybernetics, 19 (3) pp. 462-472

Kock, Ned (2013): Action Research: Its Nature and Relationship to Human-Computer Interaction. In: Soegaard, Mads and Dam, Rikke Friis (eds.). "The Encyclopedia of Human-Computer Interaction, 2nd Ed". Aarhus, Denmark: The Interaction Design Foundation. Available online at

Kvale, Steinar and Brinkmann, Svend (2009): InterViews: Learning the Craft of Qualitative Research Interviewing. Sage

Lazar, Jonathan, Feng, Jinjuan Heidi and Hochheiser, Harry (2010): Research Methods in Human-Computer Interaction. Wiley

Legard, Robin, Keegan, Jill and Ward, Kit (2003): In-depth interviews. In: Ritchie, Jane and Lewis, Jane (eds.). "Qualitative research practice: a guide for social science students and researchers". London, United Kingdom: Sagepp. 138-169

Lehn, Dirk vom and Heath, Christian (2005): Accounting for new technology in museum exhibitions. In International Journal of Arts Management, 7 (3) pp. 11-21

Lincoln, Yvonna S. and Guba, Egon G. (1985): Naturalistic Inquiry. Thousand Oaks, USA, Sage

Lindtner, Silvia, Chen, Judy, Hayes, Gillian R. and Dourish, Paul (2011): Towards a framework of publics: Re-encountering media sharing and its user. In ACM Transactions on Computer-Human Interaction (TOCHI), 18 (2)

Lipson, Juliene G. (1997): The politics of publishing: protecting participants confidentiality. In: Morse, Janice M. (ed.). "Completing a qualitative project: details and dialogue". Thousand Oaks, USA: Sage

LÃ¼tzhÃ¶ft, Margareta, Nyce, James M. and Petersen, Erik S. (2010): Epistemology in ethnography: assessing the quality of knowledge in human factors research. In Theoretical Issues in Ergonomics Science, 11 (6) pp. 532-545

Mackay, Wendy E. (1999): Is Paper Safer? The Role of Paper Flight Strips in Air Traffic Control. In ACM Computing Surveys (CSUR), 6 (4) pp. 311-340

*Mackay, Wendy E. and Fayard, Anne-Laure (1997): HCI, Natural Science and Design: A Framework for Triangulation Across Disciplines. In: Proceedings of DIS97: Designing Interactive Systems: Processes, Practices, Methods, & Techniques 1997. pp. 223-234*

*Makri, Stephann and Blandford, Ann (2012): Coming across information serendipitously – Part 1: A process model. In Journal of Documentation, 68 (5) pp. .684-705*

*Makri, Stephann, Blandford, Ann, Gow, Jeremy, Rimmer, Jon, Warwick, Claire and Buchanan, George (2007): A library or just another information resource? A case study of users' mental models of traditional and digital libraries. In JASIST - Journal of the American Society for Information Science and Technology, 58 (3) pp. 433-445*

*Makri, Stephann, Blandford, Ann and Cox, Anna L. (2008): Using Information Behaviors to Evaluate the Functionality and Usability of Electronic Resources: From Ellis's Model to Evaluation. In JASIST - Journal of the American Society for Information Science and Technology, 59 (14) pp. 2244-2267*

*Marshall, Martin N. (1996): Sampling for qualitative research. In Family practice, 13 (6) pp. 522-526*

*Mays, Nicholas and Pope, Catherine (2000): Qualitative research in health care: Assessing quality in qualitative research. In BMJ: British Medical Journal, 320 (7226)*

*McKechnie, Lynne E., Serantes, Lucia C. and Hoffman, Cameron (2012): Dancing around the edges: the use of postmodern approaches in information behaviour research as evident in the published proceedings of the biennial ISIC conferences. In: Proceedings of ISIC the information behaviour conference 2012 September 4-7, 2012, Tokyo, Japan.*

*Mentis, Helena M., Reddy, Madhu and Rosson, Mary B. (2013): Concealment of Emotion in an Emergency Room: Expanding Design for Emotion Awareness. In Computer Supported Cooperative Work (CSCW), 22 (1) pp. 33-63*

*Miles, Matthew B. and Huberman, Michael A. (1994): Qualitative Data Analysis: An Expanded Sourcebook. Sage*

*Millen, David R. (2000): Rapid Ethnography: Time Deepening Strategies for HCI Field Research. In: Proceedings of DIS00: Designing Interactive Systems: Processes, Practices, Methods, & Techniques 2000. pp. 280-286*

*Morse, Janice M. (1997): Completing a qualitative project: details and dialogue. Thousand Oaks, USA, Sage*

*Nazroo, James and Arthur, S. (2003): Designing fieldwork strategies and materials. In: Ritchie, Jane and Lewis, Jane (eds.). "Qualitative research practice: a guide for social science students and researchers". London, United Kingdom: Sagepp. 109-137*

*Norgaard, Mie and Hornbaek, Kasper (2006): What do usability evaluators do in practice?: an explorative study of think-aloud testing. In: Proceedings of DIS06: Designing Interactive Systems: Processes, Practices, Methods, & Techniques 2006. pp. 209-218*

*O'Kane, Aisling A. and Mentis, Helena (2012): Sharing medical data vs. health knowledge in chronic illness care. In: Proceedings of the 2012 ACM annual conference extended abstracts on Human Factors in Computing Systems Extended Abstracts May 5-10, 2012, Austin, USA. pp. 2417-2422*

*Odom, William, Harper, Richard, Sellen, Abigail, Kirk, David and Banks, Richard (2010): Passing on & putting to rest: understanding bereavement in the context of interactive technologies. In: Proceedings of ACM CHI 2010 Conference on Human Factors in Computing Systems 2010. pp. 1831-1840*

*O"Connor, Liam (2011). Workarounds in accident and emergency & intensive therapy departments: resilience, creation and consequences. University College London http://www.ucl.ac.uk/uclic/studying/taught-courses/distinction-projects/2009_theses/OConnorL.pdf*

*Palen, Leysia (1999): Social, individual and technological issues for groupware calendar systems. In: Proceedings of the SIGCHI conference on Human factors in computing systems May 15-20, 1999, Pittsburgh, USA. pp. 17-24*

*Pennathur, Priyadarshini R., Thompson, David, III, James H. Abernathy, Martinez, Elizabeth A., Pronovost, Peter J., Marsteller, Jill A., Gurses, Ayse P., Lubomski, Lisa H. and Kim, G.R. (2013): Technologies in the wild (TiW): human factors implications for patient safety in the cardiovascular operating room. In Ergonomics, 56 (2) pp. 205-219*

*Perera, M. (2006). Human error in context: a study of post completion error in chip and pin transactions.*

*Rajkomar, Atish and Blandford, Ann (2012): Understanding infusion administration in the ICU through Distributed Cognition. In Journal of Biomedical Informatics, 45 (3) pp. 580-590*

*Rajkomar, Atish, Blandford, Ann and Mayer, Astrid (2013): Coping with complexity in home hemodialysis: a fresh perspective on time as a medium of Distributed Cognition. In Cognition, Technology & Work,* Randall, Dave and Rouncefield, Mark (2013). Ethnography. Retrieved 15 January 2014 from [URL to be defined - in press]

Randell, Rebecca, Ruddle, Roy A., Mello-Thoms, Claudia, Thomas, Rhys G., Quirke, Phil and Treanor, Darren (2013): Virtual reality microscope versus conventional microscope regarding time to diagnosis: an experimental study. In Histopathology, 62 (2) pp. 351-358

Reddy, Madhu and Dourish, Paul (2002): A finger on the pulse: temporal rhythms and information seeking in medical work. In: Churchill, Elizabeth F., McCarthy, Joe, Neuwirth, Christine and Rodden, Tom (eds.) Proceedings of the 2002 ACM conference on Computer supported cooperative work November 16 - 20, 2002, New Orleans, Louisiana, USA. pp. 344-353

Rochlin, Gene I. (1999): Safe operation as a social construct. In Ergonomics, 42 (11) pp. 1549-1560

Rode, Jennifer A. (2011): Reflexivity in digital anthropology. In: Proceedings of ACM CHI 2011 Conference on Human Factors in Computing Systems 2011. pp. 123-132

Rode, Jennifer A., Toye, Eleanor F. and Blackwell, Alan (2004): The fuzzy felt ethnography-understanding the programming patterns of domestic appliances. In Personal and Ubiquitous Computing, 8 (3) pp. 161-176

Roethlisberger, Fritz J. and Dickson, William J. (1939): Management and the Worker. Cambridge, USA, Harvard University Press

Rogers, Yvonne (2012): HCI Theory: Classical, Modern, and Contemporary. Morgan and Claypool

Rogers, Yvonne, Sharp, Helen and Preece, Jenny (2011): Interaction Design. Wiley

Sanderson, Penelope and Fisher, Carolanne (1994): Introduction to This Special Issue on Exploratory Sequential Data Analysis. In Human-Computer Interaction, 9 (3) pp. 247-250

Skeels, Meredith M. and Grudin, Jonathan (2009): When social networks cross boundaries: a case study of workplace use of facebook and linkedin. In: GROUP09 - International Conference on Supporting Group Work 2009. pp. 95-104

Smith, Jonathan A. (2008): Qualitative psychology: a practical guide to research methods. Sage

Smith, Penn, Blandford, Ann and Back, Jonathan (2009): Questioning, exploring, narrating and playing in the control room to maintain system safety. In Cognition Technology and Work, 11 (4) pp. 279-291

*Stawarz, Katarzyna M. (2012). An Ergonomic Evaluation of the Potential Impact of Touch-Screen Tablets on Office Workers. http://www.ucl.ac.uk/uclic/studying/taught-courses/distinction-projects/2011_theses/Stawarz_2011*

*Suchman, Lucy A. (1987): Plans and Situated Actions: The Problem of Human-Computer Communication. New York, Cambridge University Press*

*Thimbleby, Harold (2008): Write now. In: Cairns, Paul and Cox, Anna L. (eds.). "Research Methods in Human-Computer Interaction". Cambridge University Presspp. 196-211*

*Wenger, Etienne (1999): Communities of Practice: Learning, Meaning, and Identity. Cambridge University Press*

*Willig, Carla (2008): Introducing qualitative research in psychology. Open University Press*

*Wixon, Dennis (2003): Evaluating usability methods: why the current literature fails the practitioner. In Interactions, 10 (4) pp. 28-34*

*Wolcott, Harry F. (2009): Writing Up Qualitative Research. Sage*

*Wong, William B. L. and Blandford, Ann (2002): Analysing Ambulance Dispatcher Decision Making: Trialling Emergent Themes Analysis. In: Proceedings of the HF2002 Human Factors Conference November 25-27, 2002, Melbourne, Australia.*

*Woolrych, Alan, Hornbï¿½k, Kasper, Frï¿½kjï¿½r, Erik and Cockton, Gilbert (2011): Ingredients and Meals Rather Than Recipes: a Proposal for Research That Does Not Treat Usability Evaluation Methods As Indivisible Wholes. In International Journal of Human-Computer Interaction, 27 (10) pp. 940-970*

*Yardley, Lucy (2000): Dilemmas in qualitative health research. In Psychology and health, 15 (2) pp. 215-228*

## About the author

### Ann Blandford



Has also published under the name of:
"A. Blandford" and "A. E. Blandford"

Personal Homepage:
ucl.ac.uk/uclic/people/a_blandford/

Current place of employment:
University College London

Ann Blandford is Professor of Human-Computer Interaction in the Department of Computer Science at University College London, and served as Director of UCL Interaction Centre (UCLIC) (2004-2011). Her teaching includes User-Centred

Evaluation Methods on the MSc in HCI with Ergonomics at UCL. She started her career in industry, as a software engineer, but soon moved into academia, where she developed a focus on the use and usability of computer systems. Ann leads research projects on human error and on interacting with information, with a focus on modelling situated interactions. In particular, she leads an EPSRC Platform Grant on Interactive Systems in Healthcare, and an EPSRC Programme Grant, CHI+MED, on Human-Computer Interaction for Medical Devices. She has been technical programme chair for several conferences, the most recent being NordiCHI 2010. See http://www.ucl.ac.uk/uclic/people/a_blandford/ for more detail.

## Your notes and thoughts on chapter 52

**NOTES:**