From Quality in Use to Value in the World

Gilbert Cockton
School of Computing & Technology,
University of Sunderland,
PO Box 299, Sunderland SR6 0YN, UK
Gilbert.Cockton@sunderland.ac.uk

ABSTRACT
This paper argues that a focus on quality in use limits the potential of HCI. It summarizes how novel approaches such as Grounded Design can let us go beyond usability to reveal the fit between designs and expected contexts of use. This however is still not enough. It cannot resolve dilemmas about what is and is not a usability problem, or when fit is or is not essential. Such dilemmas can only be resolved by an understanding of the *value* that artifacts aim to deliver in the world. HCI must move beyond contextual description to prescriptive approaches to value in the world.

Author Keywords
Keywords: Design, Quality, Fit, Value.

ACM Classification Keywords
ACM: H.1.2 – User/Machine Systems

INTRODUCTION

This paper presents an argument. Firstly, the move within HCI from simple generic usability to complex contextual fit is argued to be inadequate. The limitations of this move are shown in a novel approach that allows direct assessment of fit for interactive systems, but still, like usability approaches, fails to address and answer what should be HCI's dominant question: what value in the world is worth seeking, and how do we specify, realize and measure it?

This paper rides roughshod over the literature and professional practice in order to focus the community on HCI's one last challenge. A fair treatment of HCI's progress needs a review of research and practice that can't fit into four pages. Rather than pretend that this can be done, the paper sticks to arguing a position. The question is less whether we can "disprove" the position, but whether we should. Should we want to languish in a slippery ephemeral world of "it depends" or would we rather move to where design and evaluation are driven by clear goals grounded in solid convictions about product value?

A HIERARCHY OF WORTH FOR DESIGN

Quality in use is often seen as *the* core HCI goal. However, when usability is seen as an intrinsic source of value, we commit to absolute definitions of concepts such as "usability problems", i.e., that we can define usability in terms of abstract and generic constructs which ignore the purpose of systems and their fit to intended contexts of use.

Absolute universal definitions are the tip of an iceberg of confusion. They drive us down into essentialism, and a belief that artifacts can *in themselves* be usable. This quickly appears foolish when applied to the physical world. Interactive artifacts are no more intrinsically "usable" than physical ones are intrinsically "strong". When we say that a chair is strong, we express expectations about it *in specific contexts.* Thus a chair is strong if it can bear the weight of any expected sitter (or stander) over a reasonable life time. Physical strength can only be defined with respect to behaviors in generalized scenarios.

Usability is no different. Guidelines that recommend "good" features and reject "bad" ones cannot deliver it. Nor can we look to a definition of "usability problem" (ignoring just how confused and contradictory these definitions are [6]) to determine what should and should not count as a usability problem in a user test. Usability cannot reside in features or generic definitions. It arises in use in specific contexts, to which usability is always relative, just as the "strength" of a chair is relative to usage.

This contextual nature of usability can be implicitly recognized via "severity" (but is often not: universal severity scales do exist). User difficulties are severe if they lead to unacceptable task performance or costs. However, what is "unacceptable" depends on usage contexts. Some task failures may be acceptable. A usability approach may wonder why such tasks are supported, but what matters more are tasks that are not supported but should be!

Missing functionality cannot be unusable, no more than a non-existant chair can be weak. Quality in use restricts us to user experience with a system's implemented functionality. During usability inspection, an analyst may note the absence of support for key tasks or users, but this is not a "prediction" that can be "validated" in user tests. Instead, we are thrown onto reasoning about what is and isn't a key task. The response is analytical, not empirical.

Fit Matters More Than Quality in Use

"User needs", "key tasks", "acceptable quality and cost" can only be defined relative to usage. Nothing within systems or interactions can directly or authoratively indicate needs, criticality or acceptability. These can only be determined

CHI 2004, April 24–29, 2004, Vienna, Austria.
ACM 1-58113-703-6/04/0004.

directly and with authority via contextual research, which may combine observation, interviews, artifact analysis and other social research methods.

In the hierarchy of design worth, usability comes below fit because it is necessary, but not sufficient, for worthwhile design. Quality in use can only be achieved for implemented functions and interaction. Misfit arises as much from what is missing as from what is there. Basic quality in use approaches can't put anything back. Evaluation thus must focus on demonstrating fit as well as measuring quality in use. We must know before user testing what counts as "good fit" and design tests to assess it. For example, if a web-site is intended for international travelers, then we should expect some multi-lingual support. If this exists, then testing must involve a range of users who can demonstrate that this support is effective.

Real HCI people know all this: guidelines don't work; usability is contextual; systems must be tested in realistic contexts; usability requirements must be set before design or evaluation. But these get continually ignored: someone tries to design universally good features (later generalized to guidelines); some define usability problems in terms of "product defects"; novel interaction methods get "tested" in artificial user studies without grounded evaluation criteria (aiming "to see what happens" rarely fails!)

The smart money is 100% on contextual HCI. So why do we keep slipping back into absolute usability, the magic in cool design like strength in an indestructible chair? It's because we have no way yet to systematically demonstrate fit between system and usage contexts. We can take quantitative measures during testing and collect novels-worths of utterances and behaviors in ethnographic studies. However, it's rarely clear which quantitative measures matter [7], nor are the design implications of contextual data unequivocal (or even detectable).

Over the last eight years, I have sought ways to create better links between context, design and evaluation [4]. Within a few years, I realized that these "links" could all be understood as "fit", providing an opportunity to recover the certainties of physical ergonomics. In this Newtonian paradise, we accept immediately once shown that a pole for standing passengers on public transport is too thin and slippery for an elderly grip [1]. We do not vanish into a mire of "it depends" and "what ifs". If we could "see fit" for intangible interactions beyond Newton's world, then we could start living up to contextual values. The next section shows that we can "see fit" for intangible interaction. However, to complete the "hierarchy of worth for design", we will still need to take one further step up.

Value Matters More Than Fit

The next section uses a central heating controller as an example. In the UK, these generally control when the heating boiler (furnace) is switched on and off. They are programmed to activate the boiler during specific periods. We use them to save money by only heating rooms or water as necessary. Were energy sources free (and inexhaustible), then no-one would want one. I have no intrinsic desire to program my central heating controller. Rather I do not want to waste money (or go into debt or waste finite fossil fuels). Quality in use approaches to controller programming would *completely* miss the point. Even if such programmers are easy to learn and use, they fail to deliver full value unless they maximize savings. I know of no central heating controller that tells me how much money (and fuel) I've saved by setting a particular program. So, controllers may be designed to be easy to use and learn, but they can achieve that *without delivering the full value that we really seek.* As will be seen, fit itself is necessary, but not sufficient. Misfit, like poor usability, can only be assessed relative to a product's intended value. Without value criteria, all instances of misfit and poor usability have similar priority and severity.

HCI is Upside Down

In the hierarchy of worth for all design, quality in use is at the bottom, fit in the middle and value at the top. Both quality and fit depend on value, and quality depends on fit. Low level achievements have limited worth in the absence of high level ones. *The most important goal is to achieve value.* Quality in use matters because poor usability can and does destroy value. It cannot however create it. Trivial and worthless systems could be easy to use. Nor does basic quality in use achieve fit to usage contexts. It cannot restore missing features such as the inability of central heating controllers to predict fuel costs. Missing features matter when they create misfit between a system and its usage context. *How much* they matter depends, just as with quality in use, on how much system value is destroyed. I don't know the cost of a program until my utility bill arrives, but I do know that turning the heating off can save money. I thus don't lose all of the value of a heating controller by having no cost predictions. However people with a rigid low budget do: some may overspend; some will rarely use heating, resulting in hypothermia.

HCI has concentrated on the least important property of interaction. It has to move up the hierarchy of design worth, through fit and onto value. The next section presents a method to do better on the way.

SEEING FIT FOR HOME CONTROL

When considering users with a tight budget, an environmental conscience or both, how do we assess fit to their needs? A "good" controller only enables heating as needed, which depends on home insulation, external temperature and occupancy. If it's warm outside, no heating is needed. Houses with no people, orchids or tropical fish inside can get fairly chilly before needing heat. If all water systems are drained, heating can always be off.

Fit reduces to comparing heating needs with program capabilities. To understand needs, we must understand household behavior. Several research methods can produce relevant data here: participant observation, interviews, surveys and automatically logging the operation of the heating system. Data could be analyzed to segment users into different groups by lifestyles (and hence heating needs). Needs could be readily represented by a calendar colored to represent the need for no, minimal (frost free), moderate or high heat. Similarly, given a controller specification, weather data and insulation data on a home, we could automatically generate calendars to illustrate the consequences of different program settings. Figure 1 shows idealized 5 day calendars (hours run left to right).

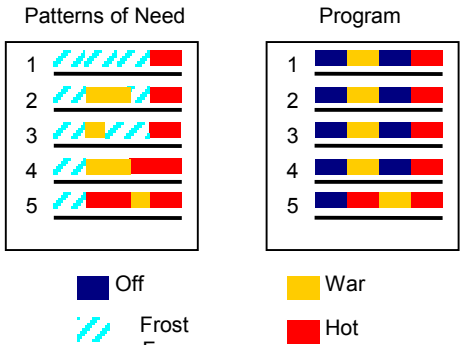

Figure 1: Needs and Realities for Central Heating

Figure 1 lets us see fit because the representations used for needs and realities have the same format, allowing ready and direct comparison. A stereotypical calendar (left) represents a user group's needs. It can be compared with a program (right) to establish the degree of misfit. Clearly, what these users need and what the controller can deliver are poorly matched. This gets us thinking about why the misfit arises, and in turn about how to design controllers with no misfit. User testing alone would never get us to this position, and nor would ethnographic studies that merely *describe* home needs. We must be able to see how designs and contexts *rendezvous* during usage.

Sources of Misfit

Central heating programmers expect us to predict our lives as regular patterns. Only some separate weekends from weekdays; others have more flexible concepts of working and home days. Others allow separate daily programs. Increased flexibility however requires more user effort, and can erode value through reduced quality in use. A simple 'fit the system to the task' approach breaks down here if the task is seen as programming a central heating controller (and that is how many device-dependent task analysis methods *would* view the task!) Fit could be achieved by adding complexity with no increase in value.

A Note on Grounded Design

Grounded Design aims to create certainties of fit as in physical ergonomics. It establishes a framework within which design fit can be rigorously and reliably assessed despite the intangibility of interaction. It takes its name from *Grounded Theory* [9], a qualitative social research methodology that systematically relates induced theories to raw data. Grounded Design seeks to ground designs rather than theories, requiring representations of interactions and outcomes that can be reliably related to systems and usages.

Grounded Design uses paired representations as in Figure 1 (called *rendezvous representations*) to respect the intellectual values of both human and computer sciences. Qualitative methods relate rendezvous representations to context, delivering the rigor sought in Grounded Theory and elsewhere. The rigor expected for computer science come from *calculating* system rendezvous representations.

Grounded Design will require extensive tool support for grounding, calculating and comparing rendezvous representations. Success depends on selecting metaphors that can be readily translated into visualizations. A series of metaphors has been developed to aid visual comparison of rendezvous representations. The first was an arch metaphor with rendezvous at *keystone* representations [2]. This couldn't cope with multiple comparisons. A gothic cathedral ceiling metaphor allowed rendezvous of ribs rising from multiple columns to meet at ceiling *bosses*. This proved to be too complicated to work with.

Recently, a metaphor of *nested shells* has simplified visualization. A system core nests within two outer shells. The first's inner surface surrounds the core; its outer surface comprises system rendezvous representations of interactive behaviors and outcomes. The outermost shell's outer surface contacts the embracing context of use; context rendezvous representations form its inner surface. We can see fit between a system and its usage context by comparing abutting surfaces of outer shells. For perfect fit there are no gaps. In poor fit, adjacent surfaces differ, resulting in gaps. Figure 2 shows cross sections. The system is the white core and the context is the embracing square. Context (blue hashed) and system (red dotted) rendezvous shells surround the core. In a design tool, the abutting shell surfaces would be 'tiled' with rendezvous representations.

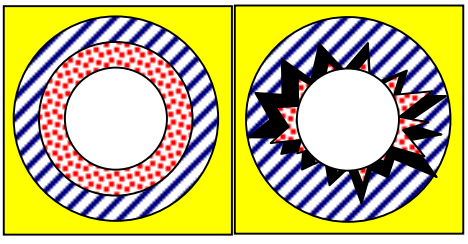

Figure 2: Perfect and Poor Fit

The left of Figure 2 shows perfect fit, with identical rendezvous surfaces. The right shows poor fit. Unaddressed aspects of context result in black gaps. System features relate unevenly to the usage context. A visual design tool based on such shells would task HCI teams with creating rendezvous representations (e.g., Figure 1) to cover all interactions between systems and usage contexts. In a perfect world, an HCI team could ground, calculate and compare all representations necessary to ensure full value with absolute quality. However, such marvelous tools would still lack representations of value.

VALUE: THE FINAL CHALLENGE FOR HCI

For over a decade, HCI has aimed to leave the lab for the real world. With the demise of absolute universal metrics [10], usage contexts became the basis for achieving quality in use. We had moved from looking for usability *in the product*, to looking for needs, behaviors and capabilities in human contexts. That is indeed where fit lies, but it is not where value is created. The usage context only delivers fit. Like usability, misfit matters when it destroys system value. While misfit may be seen via rendezvous representations, its *assessment* depends on intended value.

Value is created in the intentions of system developers. Well-designed products and services have clear *value propositions* that express the collective intentions of all those who shape them, from marketing through to design and testing. However, value tends to be poorly expressed during development. Intended value can be most clear in product proposals and marketing collateral, but too often final products disappoint. Conversely, a digital product may be a market success by delivering unintended or unimagined value (SMS, or sending text messages from mobile phones, is an excellent example).

WHERE NEXT: EMOTION OR ECONOMICS?

Grounded Design lets us pose a critical HCI question: at what point does a loss of fit and/or quality in use entail a loss of value? However, it does not let us answer this question in its current form. Like so many digital products, it does not deliver what was intended! However, its approach may be modifiable to explore delivery of value in HCI. Grounded Design could bring value as well as fit and quality into scope, allowing overall assessment of *useworthiness*. It could do so in a *disciplined* manner, that is, in a public, objective and repeatable form.

To address value, Grounded Design needs to move from the present to the future. It needs to go beyond visual comparison of rendezvous representations to being able to *calculate* the value that systems can realize. This will require contextual research to move beyond users, goals and activities to sources of value and their measurement (e.g., cost and fuel savings from central heating programming, percentage of great TV programs that I miss but get to watch within a few weeks of broadcast).

It is not clear how to move contextual research from the present to the future, from description to innovation, from current practices to future gains. Economics may provide some of the answers: there are a few inspiring examples of applying economic concepts in HCI [5, 8]. Affective computing and beyond may provide other answers [3]. What is clear is that it is time for HCI to move on, before researchers in better placed disciplines relegate us to overseeing other people's value tests.

REFERENCES

[1] Brooks, B.M., "Bus Design: A study of passenger capabilities and requirements, " in *Design for Work and Use*, Vol. 2., eds. H. G. Maule and J. S. Weiner, 15-48, Taylor and Francis, 1981.

[2] Cockton, G., "L'Avenir de L'Interface — The Future of the Interface," in *Actes des 2èmes journées Multimédia*, ed. M. Noirhomme-Fraiture, Presses Universitaires de Namur, Belgium, 11-34, 2002.

[3] Cockton, G., "Editorial: From doing to being: bringing emotion into interaction," *Interacting with Computers*, 14(2), 89–92, 2002.

[4] Cockton, G., Clarke, S., Gray, P.D., and Johnson, C., "Literate Development: Weaving Human Context into Design Specifications," in *Critical Issues in User Interface Systems Engineering*, eds. D. Benyon and P. Palanque, Springer Verlag: London, 227-248, 1996.

[5] Johnson, C.W., "The Impact of Utility and Time on Distributed Information Retrieval," in *People and Computers XII,* eds.H. Thimbleby, B. O'Conaill and P. Thomas, Springer Verlag, 191-204, 1997

[6] Lavery, D., G. Cockton & Atkinson, M., "Comparison of evaluation methods using structured usability problem reports," in *Behaviour and Information Technology,* 16(4/5), 246-266, 1997.

[7] Newman, W.M., & Taylor, A.S., "Towards a methodology employing critical parameters to deliver performance improvements in interactive systems," in *Proceedings of INTERACT 99*, eds. A. Sasse and C. Johnson, 606-612, IOS Press, 1999.

[8] Riegelsberger, J., Sasse M. A., and McCarthy J. D.: "Trust at First Sight? A Test of Users' Ability to Identify Trustworthy e-Commerce Sites," in *People and Computers XVII*, eds. E. O'Neill, P. Johnson and P. Palanque, Springer-Verlag, 243-259, 2003.

[9] Strauss, A., and Corbin, J., *Basics of Qualitative Research: Grounded Theory Procedures and Techniques*, SAGE, 1990.

[10] Whiteside, J., Bennett, J., & Holtzblatt, K., "Usability engineering: Our experience and evolution," in *Handbook of Human-Computer Interaction*, ed. Helander, M., North-Holland, 791-817, 1988.