A Load of Cobbler's Children: Beyond the Model Designing Processor

Gilbert Cockton
Faculty of Arts, Design and Social Science, Northumbria University,
Newcastle upon Tyne, NE1 8ST,UK
Gilbert.Cockton@northumbria.ac.uk

CHI'13, April 27 – May 2, 2013, Paris, France.

Abstract
HCI has developed rich understandings of people at work and at play with technology: *most* people that is, except designers, who remain locked in the information processing paradigm of first wave HCI. Design methods are validated as if they were computer programs, expected to produce the same results on a range of architectures and hardware. Unfortunately, designers are people, and thus interfere substantially (generally to good effects) with the 'code' of design methods. We need to rethink the evaluation and design of design and evaluation methods in HCI. A logocentric proposal based on resource function vocabularies is presented.

Author Keywords
Design methods; validation; resources; key words.

ACM Classification Keywords
H.5.0. Information interfaces and presentation (e.g., HCI): General

General Terms
Design, Human Factors, Theory.

Introduction
HCI research has several main foci, including the nature of human interaction with computers, the quality of innovative interaction designs, and practices that

should result in high quality designs and interactions. Much research on the latter focuses on support for design work via methods, tools, techniques, guidelines, heuristics and other forms of design knowledge. As with any claimed research contribution, these must be assessed in some way. This could relate to the impact on the design process, the resulting designed artifact, or both: one or both should be better as a result of using innovative design practices from HCI research.

Assessing research in support of design work thus logically requires some notion of what *better* means in the context of Interaction Design. However, other evaluation criteria are often introduced that do not relate to improved *design* processes or outcomes. The next section describes common requirements for HCI research in support of interaction design work.

How is this a Good Method?

The goodness of interaction design practice innovations within HCI is assessed in many ways in my experience. Some of this is based on published literature (e.g. [15]), but far more is based on 15 years of experience of relevant reviews as an author, journal editor, and conference programme committee member. Too often, papers are rejected for reasons that have no basis in scholarly values, but instead are the result of personal subjective confusions that have never been exposed to public discussion within the broader HCI community.

This paper makes this discussion public, and proposes a novel position on ways to understand, assess and create design and evaluation methods. It first examines a range of common positions on key evaluation criteria, that mis-assess isolated methods in terms of originating rationales, usage accuracy, causal roles, the easily measured, and the best of breed, often brushing major methodological challenges under the carpet.

What Problem Is Being Fixed?

The core position here is that design and evaluation methods, like all designed artefacts, must be a response to known user needs: a human-centred design position. Alternatives are possible and the HCI community should be open to taking them seriously. Verganti's Design Driven Innovation [20] rejects the need for primary research evidence of user needs and wants. Instead, his primary research evidence focuses on organizations' successful design innovation irrespective of the underlying design paradigm. Verganti has identified several design-led organisations (e.g., Apple, Alessi) where market success is not the result of user-centred methodologies. Verganti quotes the chairman of Artemide, an innovative lighting company: "Market? What market? We do not look at market needs. We make proposals to people…"

While possibly heresy to the human-centred, this position is well established in the millennia old Applied Arts design paradigm [7], where design purpose comes primarily from design teams and not primary human research data. Similarly, in the Engineering Design paradigm, design purpose may be expressed in terms of improving an existing artefact, rather than users needing or wanting these improvements.

Design is thus not just about problem solving, and even when it is, problems do not need to be grounded in primary user data. While CHI is an advocate for human-centred design, this conference community has never validated the universal superiority of its preferred design paradigm. Of course, we all have anecdotes and

evidence of creative or technically driven design failing due to inadequate consideration of human factors, but creative and technological colleagues can just as readily provide evidence of human-centred projects that have failed creatively or technically.

Human-Centred Problem Solving is one design philosophy. It is not the only one, nor does it have a monopoly on effective use of human insights in design: selective human-foci do not imply human-centredness, (strictly only the human sciences may be so centred, design must ultimately centre on designs).

Novel design and evaluation methods can thus be proposed because of creative opportunities (e.g., personas [11]), or because of technical opportunities (e.g., server log analyses, physiological sensing, eye tracking, pico projectors, tablet computing, mobile technologies). There is no need to establish a need here. What matters is the effectiveness of the innovative design practices. Their originating intellectual contexts do not determine their success.

Insistence that new methods address known empirically grouded problems allows dismissal of new methods without reference to their details.

Practitioner reviewers often reject papers on innovative methods because they see no need for them, which may tell us far more about the quality of their practice than the need for a new method. Human science reviewers often reject papers on innovative methods because they see no evidence of need for them, which may tell us far more about their understanding of the realities of design and innovation than the validity of the method. Such disqualifications before a competition starts are profoundly unreasonable, as well as grossly unscientific. Hypotheses and conjectures should never be rejected on the grounds of their origins and rationales, but only on the basis of evidence. A study may be dismissed because it is poorly designed and/or conducted, but that does not invalidate its motivating hypotheses or conjectures. Both proof and disproof must be competent.

Could People Use the Method Correctly?

The next confused evaluation criterion for design and evaluation methods is accuracy of their use. Hornbæk has identified this as one aspect of his third dogma for the assessment of evaluation methods, i.e., that usability evaluation proceeds as prescribed [15]. Here, the quality of the means takes precedence as an evaluation criterion over the quality of end results.

Accuracy of method use should not be a primary evaluation criterion. It could be useful diagnostically to understand why methods do not perform well in some circumstances, but the causal relationships have to be established here. They cannot be assumed. In fact, misuse of Heuristic Evaluation is the only possible explanation for some successful usability predictions [8] (perversely Hornbæk cites the research in [8] as an example of this dogma, rather than, as clearly intended, evidence that it is unfounded dogma).

This second confused evaluation criterion treats method use as if designers were model human information processors, accepting no input beyond the method, executing it perfectly and hence inevitably producing high quality design or evaluation results. In reality, high quality results are high quality results, regardless of how they were produced or how that was motivated.

The quality of the results of design work can and should be judged without reference to process or method rationales. Poor quality results may suggest poor process or rationales, but it does not automatically imply either. Similarly, a process that is judged to be poor in execution and origin can nevertheless result in outcomes that others, with no knowledge of process and method goals, would judge to be of high quality.

This criterion elevates some abstract Model Designing Processor above the quality of knowledge, expertise, skill, judgement and achievement of design teams. It fetishises correctness over effectiveness. Human scientist reviewers often criticise design or evaluation work because what they see as relevant research methods have not been used, or have been used poorly by academic standards. However, there is no automatic link between method quality and outcomes in design. Designing is not a garbage-in garbage-out activity. Designing is a self-monitoring, self-regulating process (i.e., a second order system), where flaws introduced via an earlier activity can be caught and corrected. The (loose) use of human science research methods in contextual research should thus not be reviewed as if it was a contribution to archival knowledge, since it is not. Instead, it is one of many inputs to design work, and should be only judged by the quality of the resulting design direction and influences.

Once again, inappropriate assessment criteria arise when aims or process has precedence over outcomes.

What is Responsible for Results?

The second aspect of Hornbæk's third dogma for the assessment of evaluation methods is that evaluation directly identifies usability problems [15]. As with the previous evaluation criterion, this elevates process over outcome. If design teams systematically get results from methods, the actual causal relations here only matter if we assume designers are Model Designing Processors, who design or evaluate on the basis of the method, the whole method, and nothing but the method. The reality is that design processes combine a mix of primary and secondary data with local resources from the design team and project setting. It is simply not possible to isolate methods from their surrounding human contexts in design work.

This is why Heuristic Evaluations can predict problems for which there is no applicable relevant heuristic within the set in use [8]. However, given that it is only methods in use that achieve anything, we can only judge methods in use. Isolating direct causal factors at method use level may be impossible, and even where this is not the case, if methods routinely bring additional benefits, then in practical terms, the method does have value, even when causal relations are indirect and even weak.

What Results Matter?

Hornbæk's first, fourth and seventh dogmas for the assessment of evaluation methods are the first to focus on the results of method use, rather than on process or rationale. Each is specific to usability evaluation, questioning the use of counts of usability problems, their use as the unit of analysis, and the assumption that they exist. Thus even where evaluation criteria are focused on results, these may not be the results of most value, but instead something that is easy to measure and count.

More Method Validation Dogmas

Hornbæk's second dogma is also specific to usability evaluation, and concerns methodological problems associated with matching difference instances of the same usability problem. This is a more subtle dogma that when generalised, allows researchers to sidestep major challenges to rigorous credible assessment.

Hornbæk's fifth and sixth dogmas are at a high level of abstraction; with the former referring to the tendency assess methods in isolation outside the context of realistic design work, and the latter referring to the search for a best method.

Summary

Hornbæk's critique of research assessments of usability evaluations [15] is a starting point for a re-evaluation of how design and evaluation methods are assessed and created in HCI research. Soon after its publication, Hornbæk collaborated with colleagues from the European MAUSE COST network (www.cost294.org) to align his critique with insights from the MAUSE project's critique of comparison studies of usability evaluation methods [9]. This formed the basis for a new perspective on evaluation methods [22] that has since evolved into a framework for understanding, assessing and improving design work. Most dogmas generalise to design and evaluation methods in general.

The key insight from Hornbæk's dogmas is that the issue of what makes design and evaluation methods *practically* worthwhile had never been in focus in HCI research. Instead, assessment of isolated methods using naïve methodologies focused on design intent (was this properly human-centred?), correct use (did designers stick to the method?), monocausality (did the method alone produce the results?), intermediate results (whatever's easiest to measure) and consumer advice (this year's best buys from the method bar).

The result of the lack of critical reflection on the conduct of design and evaluation method assessments was not only a superficial continuation of methodological flaws criticised by Gray and Salzman [13], but a deeper more insidious construction of designers and evaluators as method execution devices, *Model Designing Processors* who sought out best of breed methods for guaranteed successes. All that was needed for design could be contained within methods. Knowledge, expertise, skill, experience, and other local design team and project resources were at most confounds that the best methods would overcome, allowing identical quality performance in different interaction design settings.

A final consequence of narrow confused positions on evaluation criteria for innovative methods is that methodological innovation in HCI largely happened outside of the major research venues. For example, personas originated in practice [11], were brought from Interaction Design into HCI via the small DUX conference [18] and given rigour via a practitioners' book with extensive input from the user experience community [17] and the occasional design research contribution (e.g., [21]). It took a decade from personas dissemination by Cooper [11] for full papers on personas to appear on CHI. One of the most significant and extensive method innovations in Interaction Design had originated and remained beyond the reach of acceptable CHI research until the point where CHI publications were unlikely to influence practice. Inappropriate evaluation criteria as applied by

reviewers had kept persona and most other method research outside of the main HCI research venue.

The Cobbler's Children

"The Cobbler's Children Have No Shoes" is a observation of unknown origin that refers to the tendency of skilled workers to reserve these skills for their clients, to the neglect of the needs of themselves and their families.

Designers are the cobbler's children of HCI. While HCI has moved on substantially in breadth and depth, the richness of our understandings of users has not been extended to interaction designers. The gap here is now substantial.

Early ('first wave') HCI focused on cognitive psychology and its computationally influenced information processing model. As with much technically driven innovation, informational processing models were explored because they existed, not because there was an obvious human need for them. Computational models from Hard AI (Artificial Intelligence approaches that made claims for realism) offered new ways of understanding human-computer interaction. Card, Moran and Newell's *Model Human Processor* [3] applied a range of computationally inspired concepts to HCI.

Within a few years, the context free information processing of the Model Human Processor was challenged by Suchman's situated constructive account of technology usage [19]. Settings and social action experiences were seen as critical to understanding human interaction with computers. Over a decade later, Dourish [13] combined social interactionist perspectives with the embodied nature of human action. Our corporal and social contexts combine to produce and guide meaningful experiences with interactive technologies. A few years after this, McCarthy and Wright [10] extended user experience to include consideration of emotion and volition. Within two decades, users had become situated, embodied, affective and motivated.

Designers and evaluators in contrast remained method processors, albeit with poor reliability [14]. Working in splendid isolation, one-on-one with their method of the moment, designers' and evaluators' bodies, emotions and motivations were irrelevant. Unlike users, they had not benefitted from Third Wave HCI [1] with its ultimate embrace of all of our humanity, making cognitive information processing a dim memory as incredible as the original Charlie's Angels' hairstyles.

Enough is enough. Designers and evaluators have knowledge, expertise, skills, bodies, emotions, motivating values and social contexts. We need to understand, assess and create support for design work that exploits all Third Wave HCI perspectives when constructing models or expectations of designers and evaluators. Human Designing Processors have to be replaced with *Human Designing Explorers*, who bring their knowing, feeling expert bodies and buddies to design settings. Support for such real designers will always be integrated withing existing work practices. The most successful methods are those that disappear after first use, working invisibly and imperceptibly (to the inexpert eye) to reconfigure and empower designers and their work.

To be effective in Interaction Design, methods do not need to be:

Evaluation Resource Types [22]

1. Procedural
2. Process
3. Instrumentation
4. Expressive
5. Scoping
6. Axiological
7. Knowledge

Revised TwinTide Resource Function Vocabulary

1. Directive
2. Harvesting
3. Expressive
4. Scoping
5. Axiological
6. Knowledge

1. human-centred in origins or values
2. applied mechanically and uncritically, without local adaptation or extension
3. validated in isolation by what is easy to measure, with methodological flaws that are easy to hide
4. demonstrably superior to all alternatives all of the time

To be effective in Interaction Design, methods do not need to deterministic. We need to understand what this means, and what leads to worthwhile use.

An Alternative Approach

Hornbæk's collaborations with colleagues from the European MAUSE COST network form the basis for an alternative approach to method evaluation where methods are evaluated:

1. in realistic design settings in combination with complementary methods
2. as the achievements of design work, not as context free inert inputs to it — methods are constructed in use from 'raw' re-usable resources that need to be 'cooked' in actual design settings to become usable (edible?).

Re-usable resources may be grouped and commoditised as named approaches (branded methods), but such approaches can rarely be applied 'as is'. The vast majority require work to get them to work.

Seven *types* of resource were identified in [22] (upper list, sidebar box to left). *Procedural* resources direct design work, and constitute the core (or even entirety) of methods for many, i.e., a method is a series of steps. *Process* resources relate approaches to embracing design processes, e.g., to a specific stage such as problem analysis or design. *Instrumentation* resources collect evaluation data (e.g., logs, timers, video, eye trackers). *Expressive* resources communicate evaluation findings (e.g., problem or evaluation reports). *Scoping* resources indicate the coverage of an approach (e.g., technologies such as mobile, desktop or ambient, user groups such as children, the elderly or disabled, application domains such as games, e-learning or health information). *Axiological* resources indicate the values motivating an approach (e.g., accessibility, discount methods, value-sensitivity, situationism [2]). *Knowledge* resources underpin other resource types, and may, for example, be conceptual, factual, or relational.

The initial typology worked for evaluation methods, with one modification within the COST TwinTide project (www.twintide.org), where process resources eventually were seen as a specific form of scoping resource. However, when extended to design methods [6], some changes in vocabulary proved to be necessary to generalise over design and evaluation methods. This is shown in the bottom of the left side bar (2-6 correspond to 3-7 above and are unchanged).

These resource types were also found to correspond closely [6] to *meta-principles for designing* [5], some of which were also renamed to reduce bias towards human-centred design paradigms. *Receptiveness* [5] was judged to be too passive in a design research

workshop, and was renamed as *acquisitiveness* (which has since been associated with negative connotations in some cultures, so *Inquisitiveness* may be preferable). The original name here was Sensitivity and corresponds to the *harvesting* function. *Credibility* [5] was judged to be too positivist in a design research workshop, and was renamed as *Tenacity*.

Lastly, it was also recognised when applying an evaluation typology to design resources that they were not types, but functions, since one resource cannot have several types (complex type schemes aside), but it can have multiple functions. Sketching for example does not simply have an *expressive* function, but it also has a *harvesting* function (as an ideation technique) and also a *directive* function (in the way that sketch sequences develop through refinement, discarding and new directions). Refinement may draw on knowledge resources that guide designers in steadily improving the 'finish' of sketched elements.

Resources grouped together by approaches can be analysed in terms of their functions. This can quickly reveal gaps in approaches that have to be filled by local resources in specific design settings. Alternatively, re-usable resources can be designed to fill gaps, or complementary approaches can be identified to combine approaches to achieve coverage. It can also reveal duplication, leading to analyses of ambiguity, redundancy and complementarity, which could support simplification of approaches by removing and/or replacing superfluous or confusing resources.

Resource functions are thus a vocabulary that supports understanding, assessment and improvement of existing design and evaluation approaches, as well as targeted creation of new ones (on the basis of conceptual analysis and not demonstrated need).

Resource functions are a basis for studies of design and evaluation methods that expect approaches to interact extensively with the local resources in actual design settings. Such studies would accept alternative axiological resources to human-centred ones, approaches without directive resources (and thus no prescribed 'method' to follow), and would not expect a deterministic relationship between resource functions in their raw re-usable forms and design outcomes. Resource functions also help to explain method misuse, e.g., axiological resource functions for cultural probes can be easily overlooked [2].

Resource functions are also a basis for design teams to audit and improve their own practices. For both applications of resource function vocabularies, it is important that researchers or practitioners have a good grasp of the meaning of each function.

Taking Language Seriously

Words matter. Key words matter more. They are how we communicate and discuss the foundational concepts in any discipline. They are contentious, and evolve in disciplinary discourses.

Words are the elephants in Scientism's room, which uses technical writing to write out writing [7]. Clarity, precision and simplicity is valued, and ambiguity is to be avoided. Technical writing attempts to tame language, but may only achieve this through extensive disabling of language's finest features.

Extended Resource Function Vocabulary, with increased challenge

1. Directive
2. Inquistive
3. Expressive
4. Performative
5. Adumbrative
6. Ameliorative
7. Informative
8. Invigorative
9. Protective
10. Integrative

Everyday Resource Function Vocabulary

1. Steering
2. Sourcing
3. Recording
4. Sharing
5. Limiting
6. Valuing
7. Telling
8. Energising
9. Caring
10. Linking

The vocabulary for resource functions and related meta-principles for designing does present challenges, even for native speakers of English. Recently, it has become clear that emotional forces at work in design work introduce further resource functions. Some resources give a boost to design work, resulting in an *invigorative* function. Others keep design work on the rails, pulling design teams back from impasses and conflicts, resulting in an *protective* function. Also, complex meta-principles such as *inclusiveness* and *improvability* [5] can be unified under a single *integrative* function, with different meta-principles in scope depending on design paradigms [7]. In contrast, for expressive functions, a distinction needs to be made between expression for informal use by design creatives and *performative* expression to communicate the current state of a design's thinking to a broader group of stakeholders.

As each new function was identified, naming it became a cause for concern, as yet another cycle of renaming may be initiated. At the same time, the challenging nature of the current vocabulary was seen as a virtue. Rather than succumb to Scientism's hopeless battle with unruly language, an alternative logocentric strategy was identified, i.e., a strategy that centres on words, rather than tries to marginalise them.

Challenging neologisms have been used deliberately in psychometrics, with Cattell's 16PF traits [4] originally having names such as Protension, Autia, Parmia and Premsia. Cattell deliberately chose these terms to stop people equating his personality traits (that emerged from factor analysis) with everyday common sense terms. However, these were eventually relegated in favour of everyday English [10] (i.e., Vigilance, Abstractedness, Social Boldness, and Sensitivity). Sometimes, both vocabularies are presented in parallel.

Four experimental vocabularies for resource functions have been developed. The first existing challenging vocabulary has been made more challenging, in the spirit of Cattell. A second everyday vocabulary has been developed. The former is shown above the latter in the left side bar. A more formal but neutral technical vocabulary appears in the next side bar, with an archaic poetic vocabulary below it. The latter is deliberately exotic, from a list of historic occupations (www.rootsweb.ancestry.com/~usgwkidz/oldjobs.htm).

Rather than fix the meaning of each resource function, the vocabularies provide a space within which to explore meanings (perhaps supported by web searches). One is suited to everyday use with a broad group of stakeholders, two are suited to academic writing (one fancy, one plain) and the fourth is playful and intended to stretch the imagination.

Summary

Exclusively human-centred values, including ones from the human sciences, have resulted in inappropriate criteria for the assessment of innovative design and evaluation methods. An alternative conceptual framework focused on resource functions within resources has been developed to allow more open assessments of the influences of named approaches of re-usable resources on the quality of design outcomes.

Four experimental vocabularies have been developed to avoid the limitations of a single set of clear and simple terms. Instead, different genres of vocabularies have been developed to explore creative uses of writing

Technical Resource Function Vocabulary,

1. Instruction
2. Investigation
3. Registration
4. Presentation
5. Consideration
6. Prioritisation
7. Education
8. Animation
9. Remediation
10. Co-ordination

Archaic Poetic Resource Function Vocabulary

1. Apparitor
2. Mudlark
3. Scrivener
4. Bard
5. Assay Master
6. Prediger
7. Book Holder
8. Stoker
9. Palister
10. Billy Piecer

within HCI, complementing my recent use of parody [7]. This is motivated by the need to take language seriously as the medium through which we develop fundamental understandings of interaction design work.

References

[1] Bødker, S. 2006. When second wave HCI meets third wave challenges. Proc. NordiCHI '06. ACM, 1-8.

[2] Boehner, K., Vertesi, J., Sengers, P., & Dourish, P. 2007. How HCI interprets the probes. CHI '07, 1077-86

[3] Card, S.K., Moran, T.P., Newell, A. (1983) *The Psychology of Human-Computer Interaction*, Lawrence Erlbaum Associates.

[4] Cattell, R.B. 1990. Advances in Cattellian personality theory. In L. A. Pervin (Ed.), *Handbook of personality: Theory and research.* Guildford, 101-110.

[5] Cockton, G. 2009. Getting There: Six Meta-Principles and Interaction Design. CHI 2009, 2223-32.

[6] Cockton, G. Making Designing Worth Worth Designing. CHI 2012 Workshop on Methods for Accounting for Values in Human-Centered Computing. Available at ii.tudelft.nl/ValuesInDesign/submissions/cockton.pdf

[7] Cockton, G. 2012. UCD: Critique via Parody and a Sequel. Proc. CHI 2012 Extended Abstracts. ACM, 1-10

[8] Cockton, G. and Woolrych, A., "Understanding Inspection Methods: Lessons from an Assessment of Heuristic Evaluation," in People and Computers XV, Springer-Verlag, 171-192, 2001, 1-85233-515-7.

[9] Cockton, G and Woolrych, A. 2009. Comparing UEMs: Strategies and Implementation, Final Report of COST 294 Working Group 2, in *Maturation of Usability Evaluation Methods: Retrospect and Prospect: Final Reports of COST294-MAUSE Working Groups*, available from http://141.115.28.2/cost294/upload/533.pdf

[10] Conn, S.R. and Rieke, M.L. 1994. The 16PF Fifth Edition technical manual. Champaign, IL: Institute for Personality and Ability Testing, Inc.

[11] Cooper, A. 1998. *The Inmates are Running the Asylum: Why High Tech Products Drive Us Crazy and how to Restore the Sanity*. Pearson.

[12] P. Dourish. 2004. *Where the action is*. MIT.

[13] Gray, W.D., & Salzman, M.C. 1998. Damaged merchandise? A review of experiments that compare usability evaluation methods. *Human-Computer Interaction, 13*(3), 203-261.

[14] Hertzum, M., & Jacobsen, N.E. 2001. The evaluator effect: A chilling fact about usability evaluation methods. IJHCI, 13(1), 421-443.

[15] Hornbæk, K. 2010. Dogmas in the assessment of usability evaluation methods. BIT, 29(1), 97-111

[16] McCarthy, J. and Wright, P. (2004) *Technology as Experience*, MIT Press.

[17] Pruitt, J. & Adlin, T. 2006. *The Persona Lifecycle: Keeping People in Mind Throughout Product Design*. Morgan Kaufmann

[18] Pruitt, J. and Grudin, J. 2003. Personas: practice and theory. In *Proc.* DUX '03. ACM, 1-15.

[19] Suchman, L. 1987 *Plans and situated actions*. Cambridge

[20] Verganti, R. 2009. Design-Driven Innovation: Changing the Rules of Competition by Radically Innovating What Things Mean, Harvard Business Press

[21] Wisser, F.S. and Stappers, P.J. 2007. Mind the face. In *Proc.* DPPI '07. ACM, 119-134

[22] Woolrych, A. Hornbæk, K. Frøkjær, E. and Cockton, G. 2011. Ingredients and Meals Rather Than Recipes: a Proposal for Research that Does Not Treat Usability Evaluation Methods as Indivisible Wholes, Int. Journal of HCI, 27(10), 940-970.