Rigour at a trotting pace: A story from the user-centred design of smart Internet technologies

Maurice Castro
Supriya Singh
RMIT University

RMIT Business
RMIT University
Melbourne, Victoria
Email: Maurice.castro@rmit.edu.au

Abstract
We detail the tussle between the rigour and effectiveness of user studies for the user-centred design of technologies, drawing on the authors' experience within a UCD team. This discussion is as much a challenge between UCD researchers coming from different disciplines as it is between the UCD team and the technologists.

INTRODUCTION
In this paper we detail a difference of opinion about the rigour and effectiveness of methodology within the early stages of the User-Centred Design (UCD) team of the Australian Smart Internet Technology Cooperative Research Centre (SITCRC). The cross disciplinary nature of the team brought the wider issues about rigour and proof to the surface.

In this paper, we tell of a heated discussion between the two authors at the early stages of the project. The two authors were part of the same university and the same project in the SITCRC – but Singh is a sociologist and anthropologist, while Castro is a computer scientist. The differences about rigour and effectiveness came to a head while designing personas and scenarios for the technology programs. These differences connect with debates within Human Computer Interaction. They also relate to different journeys which ended in the SITCRC.

Singh came to the Centre after conducting sociological studies of the use of technologies. These studies of use were products in themselves, and not originally linked to design and the need for inter-disciplinary communication. The audience was sociologists, anthropologists and qualitative researchers. Hence the rigour of the research was a mark of excellence within that framework. It was important that the data fit the emerging theory. This desire for approval from one’s peers was not unlike the way computer scientists want the approval of their peers for simple, elegant algorithms.

Castro saw the user studies as starting points for conversations about design. The user's perspective was going to get fleshed out iteratively as the project proceeded. The outcome was the design of technologies. Hence the personas and scenarios were going to be judged by their usefulness for the technologists.

Focusing on this one altercation dramatically illustrates a methodological problem at the heart of the UCD of new technologies. How can we be rigorous about user studies and still be effective in the design process? Often it is assumed that these differences of approach are between the UCD team on the one hand and the technologists on the other. But our case shows that it was in the UCD team that we cared most personally about these methodological issues.

BACKGROUND
The Smart Internet Technology Cooperative Research Centre (SITCRC) brought together researchers from 12 universities and corporate partners. The aim was to design smart Internet technologies which people would find useful. To this end SITCRC was divided into four technology programs and a User Environment (UE) program that intersected the technology programs.

The UCD team was in itself interdisciplinary comprising social scientists, computer scientists and systems engineers. We had a common purpose in that we thought it was important to place the users at the centre of design, right from the early stages of the designing of new technologies. However, we brought to it different methodologies and attitudes to rigour.
The protagonists:

- Singh is a sociologist and anthropologist and has spent several years studying the use of information and communication technologies within a social and cultural context. Some of this work on users has informed the early users’ perspectives in the UE program (Singh 2001a, 2001b, 2001c; Singh et al., 2001; Singh and Slegers, 1997).

- Castro is a computer scientist who has worked to connect industry with its customers. It is in this context that he developed his interest in understanding the use of technologies.

THE ALTERCATION

We were in the first year of a seven year program of study in the SITCRC. The UE program was still trying to show the technological researchers there was value in adopting the UCD approach from the beginning as opposed to the traditional approach of building the prototype and then testing it to develop it for the market. To this end, we were developing personas and scenarios in the Alan Cooper style (Cooper, 1999), to present to the technology programs. Personas and scenarios, that approximate to characters and the plot of the story, we thought were an effective way of translating user studies to design elements.

Many of the personas and scenarios were backed by previous user studies. The aim was to substantiate these personas and scenarios further through the use of expert panels and further user study.

Castro’s role was to particularly link with the Intelligent Environment (IE). He drew up a detailed scenario of the use of an intelligent mobile device within a hospital environment. The device did not exist and no similar device was available. The purpose of the scenario was to explore how such a device could be used. The scenario told the story of a fictional persona named Matt who worked as a registrar in a major city hospital. As hospitals are unfriendly environments for electronic equipment, the story outlined the physical abuses to which they could be subjected. The devices needed to work with and without networking as hospital monitoring equipment should not be subjected to some types of electromagnetic radiation and hospitals are often old buildings which block signals. Most importantly, the story related snippets from Matt’s day showing how and why he would use his device for particular activities. It also examined the personal needs of the user in terms of privacy, comfort and control.

The intent of the scenarios was to shift the focus onto the needs of users rather than the needs of the designers. Cooper identifies focusing on the needs of developers as “Mud Hut” design, when building for third parties, the developers must look at the needs of future occupiers – occupiers who are not the builder - to progress from owner builders to project homes (Cooper, 1999). “Mud Hut” design was endemic in the project before personas were introduced. At no stage were the scenarios presented or represented as completed work, in fact they were presented as a starting point for a design discussion. The technology project discussions, to this point, had been couched in terms of “what can we do”. The personas allowed these discussions to encompass “what does a user want”.

Cooper further supports Castro’s application of personas when he identifies:

"Precision, Not Accuracy

As a design tool it is more important that a persona be precise than accurate. That is, it is more important to define the persona in great specific detail than that the persona be the precisely correct one... The end result is to have a program that does the right thing...

... It matters more that the persona is expressed with sufficient precision that it cannot wiggle under the pressure of development than it does that it be the right one. (Cooper, 1999, pp129-130)."

For Cooper it was important that the developers could not twist the persona to reflect the developers’ views of what the user ought to think. By supplying detail, the developers were unable to fill in the blanks to suit themselves. Furthermore, greater harm for the project would result from not having a persona than having one that was not quite right. As a result, Castro was searching for a persona and scenario that was not false, rather than one that was true; thus he was prepared to accept deficiencies in the persona drawn from a tiny sample, anecdotal evidence and incomplete literature, because the alternative was a void that would be filled in with the developers themselves. The most important role of the persona and scenario was to shift the focus to the user and his or her problems, rather than to present a definitive completed result.

The core issues raised were related to location in a sensitive environment and this had more than satisfied the primary audience of computer scientists grappling with the questions of location and its consequences for privacy.

The persona was presented to the university’s UCD researchers. Towards the end of a meeting, the leader of the group, Singh and Castro were left mulling over the afternoon’s discussions. It is then that the issue of rigour exploded.

So how did Castro construct the personas and scenarios?

- As Singh remembers it: He had talked to his brother.

- As Castro remembers it: He had integrated his work on mobile devices, location issues and stories from his brother and others in the medical profession and written the persona and scenarios to communicate this to his
co-workers. Many of the observations were simple engineering consequences, but the effect on the user of these consequences was highlighted. As an added precaution to ensure that the situations and characters were realistic he had asked his brother - who worked in such an environment - to read, verify and make suggestions.

Essentially, Singh demanded the answer to two questions:

- How had a non-expert in the area of Medicine, managed to achieve these insights?
- How do we prove the insights are right? Why should we trust the scenario?

Adding background about Castro’s previous work justified the right to have a knowledgeable opinion. In this case, the earlier work included closely supervising a student project helping a group of neurosurgeons design a multimedia database and the questions and exposure that work provided.

The second question was deeply troubling for Castro. It is impossible to prove in the mathematical sense, assertions in relation to the human condition. Even using a weaker definition for proof — showing it to be true at some time for some sample set — is extremely difficult.

Castro came from a general science degree, specialising in computer science. As a result, his background was founded in the philosophy of physical and experimental sciences. Developing personas that required deep immersion in the problem and drawing on personal experiences to come to a conclusion was an anathema to Castro, as results should be testable to be science. The deeply personal nature, the unrepeatability and the lack of durability of the result, made the persona not a product but only a hypothesis in his opinion.

It was at this point that the issue of methodology came sharply into focus. From a computer scientist’s perspective it seems that social scientists pay a great deal of attention to methodology, and it is essentially insurance. They conduct work that is hard — sometimes impossible — to repeat, so it is vital that their approaches stand up to scrutiny. In most cases they depend on observation without the ability to isolate the components of a system. Furthermore, the systems they are studying do not contain repeatable elements, and hence validation by repetition is impossible. For a computer scientist these difficulties are rarely encountered. Computer programs can be rerun; others can write similar programs and should get similar results. Computer scientists’ methodological arguments are often short and debate is about whether the program we have written corresponds to the theory being tested.

Singh and others suggested a number of acceptable ways to verify the conclusions:

- Literature review – For the social scientists, this was an acceptable way of framing current insights within a body of knowledge. From a computer scientist’s point of view, this was problematical. Though an examination of the relevant literature would be useful, it will never prove the conclusions.
- Expert panel – Selecting a group of experts to put their imprimatur on the conclusions is also possible.

Neither of these solutions is entirely satisfactory to an experimental scientist. Although there is reputational safety in numbers, Kuhn’s theory of paradigm shifts (Kuhn 1962) should make scientists cautious of accepting the current wisdom as reality. Both the literature and the expert panel are inclined towards the status quo.

Furthermore, neither of these solutions could be achieved in the short period of time available to do the user research.

The majority of computer science research use a research approach based on formulating processes, methods or algorithms and a research method of conceptual analysis (Glass et al 2004). Castro, as a product of this background, felt:

- Many of the supposed insights are quite obvious from the engineering framework; it was only necessary to bring the effect on the user to the fore.
- The emergent issues created by the personas interacting with the technology are by definition unpredictable; these are found by experiment.
- The emergent issues of implementing the technology itself are also often unpredictable.
- A major benefit of using personas is to remind computer scientists and engineers is that they are designing for people who have different needs to the designer.
- Software development (particularly when developing new concepts) is frequently iterative, providing opportunities to place the items in the hands of the user.

Improving the scenarios and personas through larger groups would be less effective than testing paper and physical prototypes with users, showing the value of the concepts in a computer science context. The spectre of unexpected technology and user issues also drive us towards experiment.

The products of computer science and software engineering are subject to high degrees of failure. In 2000, surveys found that 28% of US projects were completed on time and on budget (Standish 2001). This represents a marked improvement on previous years. A lack of user involvement in the process is one of the key factors that the Standish Group identified as central to project failure. The personas and scenarios contribute to good specifications. In this
environment, prototyping and experiment are routine ways of reducing the risk to success and actively involving the user.

Castro agrees with Cooper (1999)) that “trials should be informed by something more than random chance, and should begin from a well thought out solution, not an overnight hack” (Cooper 1999, pp52-53).

But Castro would emphasise that only so much can be learnt from users who are unfamiliar with a proposed future vision. This learning leads to Cooper’s starting point. Going beyond that starting point requires experiment. Carefully refining the starting point may not pay dividends if the starting point proves technically infeasible or if a good enough approximation has been found.

Because the consumers of the research could verify the conclusions experimentally they may well have been satisfied with something far less rigorous, especially when there was an urgency to get the results. However, if less rigorous work is used, the consumer must be informed and evaluate the results as satisfactory for their purpose. Appropriate warning labels must be attached to the work to ensure the consumer is well informed of its limitations.

From Castro’s point of view the personas were not meant to be deep social science research. They were a means to an end, and they could only be truly verified by testing the scenarios directly with users. Singh however viewed the scenarios and personas as one of the final products from the UCD team – hence worthy of a major initial investment in rigour. Talking to one person and coming up with a scenario demeaned the painstaking sociological work relating to the use of technologies. Was there a place for sociologists in Human Computer Interaction, if scenarios had no relationship with theory? Where was the rigour in qualitative research if the negative case has not been explored? Where was the fit between data and theory? Had a diversity of users been examined? Could the work of the UCD group be trusted without this rigour?

Despite these reservations, the IE people found Castro’s scenarios useful. For a variety of unrelated reasons, they did not result in prototypes that were tested with users.

THE POST MORTEM

It took a year for the unease to mount. During that year, enough trust had developed within the UCD team to speak frankly. Significant tensions were recognised between the methods of software engineering and social research. These might be characterised as “workable solutions” vs. “discipline-based research”. But articulating this deep sense of unease was also the beginning of a continuing discussion of methodology as we worked together. As two of the proponents of these methods were at the same university, first in different faculties and later in the same faculty, we were able to resolve the tensions face-to-face and usefully illuminate the underlying issues.

Writing out the differences and the process of auto-ethnography helped place the differences within an analytic framework. Castro went back to his office and wrote a draft paper. This became a topic for further discussion. It was followed by reflection on what had happened.

Castro reflecting on the process, said:

Many of the useful observations are only discovered in the process of doing the work. This is more a characteristic of computer science than Engineering. In Engineering where there is deep body of knowledge many issues can be predicted, in Computer Science codification and tools are limited. This means that the method and approach is subject to change in the development of the project and to some extent only discoverable by conducting the project.

Singh commented on Castro:

C might be right in that technologists and social scientists approach user studies as different kinds of products.... I am still comfortable with personas when they are backed by a study of users and with scenarios when I know that the activities and context are those that one would commonly find with a variety of users. But then in the UCD literature, there is an acceptance of the need for field studies, but there is also the fact that few UCD professionals actually take the time to do it. It is as if the user studies are capital already formulated – if not one goes into “ethnography lite”. It is as if observation and fieldwork are telescoped into a day or two.

RIGOUR AND EFFECTIVENESS  IN HCI AND SOCIOLOGY

An e-mail survey of over 100 leading UCD professionals showed that Castro was more typical of UCD professionals. The survey (Mao, 2001) showed that UCD professionals ranked field studies as important, but they rarely conducted them because of their high cost. The same was true of user requirements analysis. Informal expert review was used more often than its perceived importance. This was because it was seen as low cost but having high impact.

The norm in corporate contexts is to deliver the design with six-month windows (Beyer and Holtzblatt, 1998). In the SITCRC at the discovery stage of UCD, we worked with a longer time frame, because we were working in a mixed academic and corporate context (Singh et al, 2003). But even in the SITCRC, there is tension between the social scientists’ need for rigour and the technologists’ and business partners’ desire to have more immediate outcomes (Singh et al., 2004).
Is rigour then dispensable in HCI? If HCI is to be strengthened by the contributions of social sciences, then this is a question that has to be addressed.

HCI has borrowed many techniques from sociology and anthropology. Notable among them are contextual inquiry (Beyer and Holtzblatt, 1998), ethnomethodology (Button, 2003) and a “quick and dirty” approach to ethnography (Carroll, 2003). However, the kind of ethnography practiced in HCI differs greatly from the immersion in the worlds of a group of people, that characterise the work of anthropologists and sociologists such as Geertz (1973, 1988) and Whyte (1955).

Rigour in ethnography is to go beyond talk to conversation, to convey to the reader the anthropologist’s understanding of the world from the perspective of the actors, that he or she had “been there” (Geertz, 1988, p.5). Using immersion in other worlds is as true of the ethnography of the Internet (Miller and Slater, 2000), as it is of Italian slums or Indonesian villages.

Miller and Slater (2000) in their ethnography of the Internet in Trinidad, say:

In a more narrowly methodological sense, an ethnographic approach is also one that is based on a long-term and multifaceted engagement with a social setting. In this regard we are both relatively conservative in our defence of traditional canons of ethnographic enquiry. This seems particularly important at the present time, when the term 'ethnography' has become somewhat fashionable in many disciplines. ...We assume ethnography means a long-term involvement amongst people, through a variety of methods, such that any one aspect of their lives can be properly contextualized in others....(p. 21).

In sociology, the issue of rigour has been long debated. Reliability and validity are established differently in qualitative and quantitative research. But as Morse and Richards (2002) say, it is essential “that determining reliability and validity remains the qualitative researcher's goal” (p. 168). Validity in qualitative research is not equivalent to objectivity, but refers to questions about the authenticity, trustworthiness of findings and their implications. Can people depend on them for policy or legislation (Lincoln and Guba, 2000). So what constitutes validity and reliability in rigorous qualitative research?

Morse and Richards (2002) say,

The key to rigorous qualitative inquiry is the researcher's ability to think qualitatively. This is an exhausting process of being constantly aware and constantly asking analytic questions of data, which, in turn, constantly address the questions asked. Qualitative inquiry constantly challenges assumptions, constantly challenges the obvious, reveals the hidden and the overt, the implicit and the taken for granted, and shows these in a new light (p. 170)

As the dominant thrust of qualitative research is inductive, they say, "Asking analytic questions of the data is the key to solid and significant research" (p. 172). In order to ensure the rigour of the project, we need to:

- Ask analytic questions and revisit them;
- Ensure the method chosen is appropriate;
- Have reliable collection and coding of data;
- Check relationships in the data against other studies;
- Verify the conclusions (Morse and Richards, 2002)

Ways of achieving these aims differ according to project and context. The subjects of qualitative inquiry are often those where understandings are nebulous, and the questions themselves are not known. So as the project proceeds, one of the valuable outputs is revisiting the changing question. At the end of the project, we will have an important new question as well as insights about the questions we had previously asked (Popper, 1979; Singh and Richards, 2003).

Reliability of data collection and coding is helped by the transparency of the process. One needs to elaborate what data were collected, what data were not. What were the reasons for selecting the population group? Has the process of reaching conclusions been described in a transparent way? (See Silverman, 2001). In grounded theory, qualitative researchers are offered "a set of clear guidelines from which to build explanatory frameworks that specify relationships among concepts. Grounded theory methods do not detail data collection techniques; they move each step of the analytic process toward the development, refinement and interrelation of concepts" (Charmaz, 2000, p. 510). This retains the flexibility while leading to a systematic process of analysis.

The use of computer programs for qualitative analysis help with this transparency and checking on your conclusions. This is to avoid having one memorable case dominate the analysis (Singh, 1996). The negative case analysis is important for this process, for that is where the researcher tests the conclusions. If the conclusion does not fit the negative case, then how should the conclusion be modified. Theoretical sampling is another way of further bounding conclusions for the user group or groups being studied. Triangulation is important too in this process. Here we are interpreting triangulation as "the gaining of multiple perspectives through completed studies that have been conducted on the same topic and that directly address each other's findings." (Morse and Richards, 2002, p. 76)
It is clear that anthropology and sociology have long established traditions of rigour in qualitative research. This is what defines an anthropologist and sociologist. HCI borrows from these traditions of qualitative inquiry, while modifying it for the context of design. Qualitative research in HCI remains valuable because it seeks to provide the social and cultural context of the activity and the use of technology. Without this observation and understanding of people using technology in context, a valuable dimension would be lost. Methods such as contextual inquiry are also important to give technologists an early understanding of the users’ perspective. The issue however for the social scientist remains one of rigour if he or she wants to contribute to HCI and work in the best traditions of his or her discipline.

RIGOUR AND EFFECTIVENESS IN COMPUTER SCIENCE

Computer science is a diverse discipline ranging from formal methods (used in program verification, compiler design, optimisation and cryptogprhoy) to observation (used in Human Computer Interaction and Graphics). The standard of proof, therefore, varies from that of a formal mathematical proof through to informal rules of thumb that have been observed to work but have little or no supporting theory. In Glass et al 2004 the research methods and approaches found in the published work of computer scientists were analysed. The majority of articles used a formulative approach and a conceptual analysis method. Methods not employed included grounded theory and ethnography.

Within the discipline there has been debate about appropriate standards. Many of these issues were brought to the surface when the National Academy of Science’s Computer Science and Telecommunications Board (CTSB) brought together a group of eminent computer scientists and engineers to examine the problems one group of computer scientists were experiencing within computer science. Practitioners of Experimental Computer Science and Engineering (ECSE) were finding promotion difficult within academia and part of the problem came from the judgment of their work. In CTSB (1994) the committee identified experimental computer science and the problem of judging it by the standards of maths and physics as:

*The construction of computing artifacts is not strongly coupled to theoretical computer science. Unlike the more traditional sciences (e.g., physics) in which the interplay and coupling between experiment and theory are rather tight, an "experiment" in ECSE generally does not verify a prediction from theoretical computer science or rely heavily on a model developed theoretically. The reason is that the complexity of most real computing problems precludes the direct application of analysis: a problem can be made theoretically tractable only by abstracting so extensively that the problem that emerges may not capture the essence of the original problem.*

The committee determined that the problem of evaluation had a strong influence on practice:

*The characteristics of ECSE have a major impact on how ECSE faculty must go about their work. ... The most substantive difference, perhaps, concerns how research is evaluated. Because ECSE is fundamentally a synthetic discipline, it is straightforward to create a new computational phenomenon or an alternate implementation of a concept. Yet doing so does not automatically constitute an intellectual contribution. Rather, whatever has been created must be shown to be better than some alternative. In research in theoretical computer science, the key question is, Has the proposition been proved? In contrast, the key questions in evaluating ECSE research include, Does the idea provide a new and more useful capability or greater functionality? and, Is it faster or more efficient?*

The findings of this report indicated that Computer Science as a discipline still had difficulties in judging the quality of the work of its practitioners. Furthermore, the report entreated its readers not to reject work merely because the measures were non-standard. In particular the committee used the example of the computer mouse as a significant technology, valuable in hindsight, would have been judged poorly by the standards of mathematical and scientific proof to examine the issues.

At the experimental end of the discipline, Computer Scientists have the advantage of being able to carry out experiments, some of which are repeatable. The repeatability of experiments contributes to rigour. Unlike older scientific disciplines, such as mathematics and physics, computer science has had to accept a wider range of standards of rigour. This continues to cause debate within the discipline and between disciplines.

THE RESOLUTION

Over the next two years, the approach taken to address the issues of rigour and timeliness differed according to the project, its phase and the quality of user knowledge already available.

In two projects where the UCD team was involved, the scenarios were based on rigorous field research informed by sociology and cultural theory and then modified. The Swarm scenario was based on work done for an ongoing PhD study of the use of mobile technologies by young people in Australia (Satchell, 2003). The scenario was presented to the technologists relating to the Smart Networks program of research and modified. The modified scenario was taken back three times to users over 18 months, as the design was being detailed. So the scenario benefited from the
perspectives of cultural theory, while at the same time translating cultural user requirements to technical user requirements for design.

For the Secure Identity Management Project, there was little user and activity centred research available on the control of personal information. Existing studies were from a technological or legal perspective, focused on people’s attitudes to privacy or on technologies relating to privacy. Hence we did an extensive qualitative and quantitative user study (Singh et al., 2004b). This study has yielded valuable theory about the different ways people control information about money and health. It has also started us on the relationship of user control to security, trust, privacy, and identity. However, the six month taken for the user study was one of the reasons, the user study did not feed into the design of mobile devices with privacy mechanisms. But as the themes being investigated are major issues for the design of all information and communication technologies, the investment has been at the level of the SITCRC rather than for one project alone. We have reflected on the need for rigour and timeliness, and have proposed in a separate paper a modified process of open-ended interviews, which will lead to a grounding of theory, scenarios and user requirements in the one interview (Singh et al., 2004a).

Initial scenarios for design have also been used in the Secure Identity Management Project. They were based on the experience of the researchers, and were primarily communication tools for inter-disciplinary communication in the project team. In the Amivox project which aims to design mobile devices for the blind, an initial scenario was used. It was based on the project leader’s prior experience and study of people with disabilities and their use of technologies. The scenario was then the subject of discussion with an expert panel. These processes led to the Amivox project. The scenarios and possible design suggestions were then presented to focus groups for further elaboration. Moreover, the design team also included an expert in the use of technologies by blind people. (Astbrink and Kadous, 2003).

CONCLUSION
The challenge still remains: How do UCD researchers combine rigour with rapid appraisal techniques?

It has become clear there is value in both deep and ‘lite’ approaches. However, the risks associated with ‘lite’ methods are significant for both the users and the reputations of suppliers of user research. Products in their development stage must be carefully distinguished and in the case of ‘lite’ products, consumers must be sufficiently sophisticated to recognise the deficiencies of the product and undertake to test the conclusions further in their prototyping phase. The issues of disciplinary perspectives and methodological rigour will however be with us for some time to come.

REFERENCES


Cooper, A. (1999) The inmates are running the asylum. Indianapolis: SAMS.


**ACKNOWLEDGEMENTS**

The authors gratefully acknowledge the support of our colleagues at RMIT and in the Smart Internet CRC who endured our process of discovery.

**COPYRIGHT**

Castro, M and Singh, S. © 2004. The author/s assign Griffith University a non-exclusive license to use this document for personal use provided that the article is used in full and this copyright statement is reproduced. The authors also grant a non-exclusive license to Griffith University to publish this document in full in the Conference Proceedings. Such documents may be published on the World Wide Web, CD-ROM, in printed form, and on mirror sites on the World Wide Web. Any other usage is prohibited without the express permission of the authors.