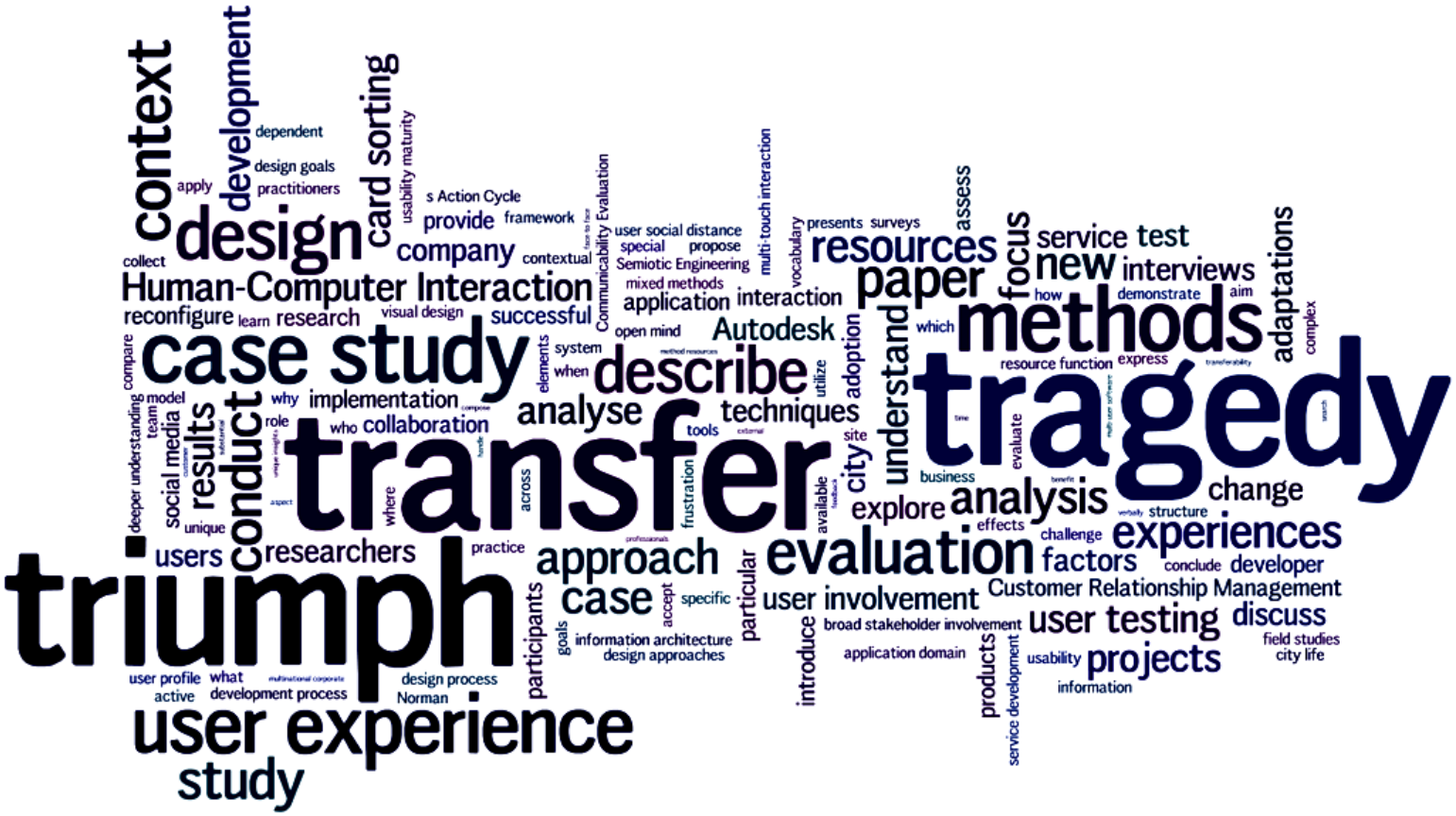
Made for sharing: HCI Stories of...

Workshop at the ACM CHI 2013 Conference



Effie Lai-Chong Law
University of Leicester, UK

Ebba Thora Hvannberg
University of Iceland

Arnold P.O.S. Vermeeren
TU Delft, the Netherlands

Gilbert Cockton
Northumbria University, UK

Timo Jokela
Aalto University, Finland

Towards the integration of transectorial IT design and evaluation

IC0904 TwinTide

Title:
Proceedings of CHI 2013 Workshop "*Made for Sharing: HCI Stories for Transfer, Triumph and Tragedy*"

Editors:

Effie L-C Law, Ebba Thora Hvannberg, Arnold, P.O.S. Vermeeren, Gilbert Cockton, and Timo Jokela

Front cover designed by: Hannah Erlebach

Printed by: Audio Visual Services (AVS), University of Leicester, UK

ACKNOWLEDGEMENTS

First of all, we are very grateful to the CHI 2013 workshop chairs, *Nadir Weibel* and *Ido Guy*, for their generous support. Thanks must also go to the authors of the workshop's papers, whose contributions serve as rich sources of stimulation and inspiration to explore the theme of the workshop from multiple perspectives. The quality of the contributions could further be ensured and improved with the generous help of the program committee members and additional reviewers. Their effective and efficient review works are highly appreciated.

List of Program Committee members:

- Arnold P.O.S Vermeeren, TU Delft, the Netherlands
- Ebba Thora Hvannberg, University of Iceland
- Effie Lai-Chong Law, University of Leicester, UK
- Gilbert Cockton, Northumbria University, UK
- Mark Springett, Middlesex University, UK
- Marta Kristín Lárusdóttir, Reykjavik University, Iceland
- Nigel Bevan, Usability Professional Services, UK
- Timo Jokela, Aalto Univesity, Finland

Last but not least, we express gratitude to our sponsor – COST (European Cooperation in the field of Scientific and Technical Research; http://cost.cordis.lu/src/home.cfm). The COST Office operated by the European Science Foundation (ESF) provides scientific, financial and administrative support to COST Actions. The COST Action IC0904 (http://www.twintide.org), which is commonly known as **TwinTide (Towards the Integration of Trans-sectorial IT Design and Evaluation)**, was officially launched in November 2009. The ultimate goal of the Action is to integrate and foster research and practice on design and evaluation methodologies for computing artefacts across sectors and disciplines. The current Workshop is the first open workshop implemented under the auspices of TwinTide. As with this and forthcoming events of TwinTide, we aim to provide the participants with enlightening environments to further deepen and broaden their expertise and experiences in the area of user experience and software development.

TABLE OF CONTENTS

HCI Stories of Transfer, Triumph, and Tragedy: An Introduction

Effie Lai-Chong Law
Department of Computer Science
University of Leicester
UK
elaw@mcs.le.ac.uk

Ebba Thora Hvannberg
School of Engineering and Natural Sciences
University of Iceland
Iceland
ebba@hi.is

Arnold P.O.S. Vermeeren
Faculty of Industrial Design Engineering
Delft University of Technology
The Netherlands
A.P.O.S.Vermeeren@tudelft.nl

Gilbert Cockton
Faculty of Arts, Design and Social Sciences
Northumbria University
UK
gilbert.cockton@northumbria.ac.uk

Timo Jokela
Usability consultant at Joticon Oy
Aalto University
Finland
timo.jokela@gmail.com

OVERVIEW

Recent studies on how traditional HCI methods are applied in practice entail re-conceptualization of the nature of such methods, leading to the notion of 'method-as-set-of-resources'. Re-usable resources provide some, but not all, of the required resources for design work. Others must be provided within design work contexts. The expanding scope of use contexts alongside the shift of emphasis to user experience calls for the development of alternative HCI practices. These two trends can influence each other. Understanding, via structured case studies, how HCI professionals transfer the same (set) of design and evaluation methods across use contexts in terms of appropriating and configuring method-resources can provide applied knowledge for: (i) creating new methods, (ii) training novices, and (iii) laying a firmer groundwork for formal analysis of HCI methods. This workshop aims to bring together HCI professionals who have method-transfer experience and knowledge to share, analyze and synthesize insights so gained.

BACKGROUND AND MOTIVATION

The shift of emphasis in the field of Human-computer Interaction (HCI) from usability to user experience (UX) challenges HCI professionals to deal with the expanding scope of interaction design. Transferring 'old' methods, with which HCI professionals are familiar, to new contexts can be a practical 'intermediate' solution. Indeed, it is observed that UX methods are largely drawn from the existing usability work ([3], [13]). However, more fit-for-purpose-and-context HCI methods should be created. Developing such methods entails a viable research programme, which can be built upon insights gained from success as well as failure cases of method transfer. Unfortunately, little is known about the process of such transfer, despite a set of related studies on usability practice in reality.

Understanding which, why, who, what, when, where and how (six W and one H questions) HCI methods are deployed by HCI professionals has been researched for about two decades; pioneer studies include [4]. More recently, two international projects networking a large group of HCI researchers and practitioners in Europe, MAUSE and TwinTide , have dealt with the six Ws and one H questions pertaining to usability evaluation methods (UEM) and to UX design and evaluation methods (UXDEM), respectively. Among different challenges tackled by the two projects, comparing UEMs and transferring UXDEM across a range of usage contexts have been seen to be fraught with difficulties.

With this workshop, we aim to gather practitioners and researchers together from the wider HCI community to examine issues on method-transfer based on the perspective of approaches and resources, and case study analyses.

WORKSHOP GOALS AND THEMES

- To collect and meta-review well-structured case studies of professional HCI practices for constructing applied knowledge for adapting and combining resources of sets of methods to deal with contextual constraints. This will be valuable for:
  - educating and training novice HCI professionals;
  - developing innovative HCI approaches to address new usage contexts;
  - laying the foundation work for formal comparisons of HCI practices;
- To deepen the understanding of how HCI professionals conceptualize HCI methods (i.e. properties, assumptions, relevance);
- To enable HCI professionals to reflect on their practice (cf. reflective practitioners [11]) by externalizing their tacit knowledge, values and strategies in relation to the roles methods play in their work in reality.

RELEVANCE TO THE FIELD

The view on the nature of HCI methods, including UEM and UXDEM, has evolved with the ongoing discussions within and outside the projects, and been crystallized in the recent publication of Woolrych and collaborators [16]. Accordingly, rather than treating a HCI method as an irreducible whole consisting of prescriptive procedural instructions, it is more appropriate to see a HCI method as a set of constituent resources, such as problem merging, heuristics, analysis, reporting formats, and task selection (for details see [16]). Only some of these resources pre-exist specific design work (often grouped into named approaches). Incomplete resources for an approach are configured and combined according to several contextual factors, scoped by project characteristics or organization where method-resources are instantiated.

As well as focusing on resources, [16] proposes a four-stage research program towards formal experimentation on resource choice and use within usability and UX work. The first and foremost stage is to capture the relevant rich context (e.g. designers' repertoire, corporate culture, product attributes) with detailed, well-structured case studies [16]). For instance, case studies on cognitive walkthrough were conducted by John and colleagues ([5]). However, wider use of this research strategy is needed.

Several studies on investigating professional usability practice have mostly employed questionnaires (e.g. [2], [8], [14]), which, albeit lacking in contextual details, provided some useful insights into which and why (and to a limited extent how) HCI methods were applied in practice. In contrast, there are only a few studies attempting to study usability work in its full dynamic complexity. Furniss, based on in-depth interviews with practitioners, built a model of how contextual factors influenced the selection and application of UEMs [7]. Similarly, Lárusdóttir and her colleagues [9] looked at how UX related activities could be integrated into Scrum projects by intensively interviewing two UX specialists. Følstad and his colleagues [6] also conducted interviews with usability professionals, though smaller in scale and scope, to understand how practitioners analyzed usability data.

No published research has focused on collecting cases that describe how practitioners transfer re-usable HCI approaches into new contexts. The process involves tacit knowledge and strategies, which will become more conscious and easier to externalize when the usage context changes [10], stimulating HCI professionals to reflect on the resources and settings for further adaptation. Such reflections and related knowledge and strategies are an integral part of a case study. Meta-reviewing a critical mass of such case studies will lead to a body of applied knowledge that is very valuable for: (i) educating and training novice HCI professionals; (ii) developing innovative HCI methods to address new usage contexts; (iii) laying the foundation work for formal comparisons of design and evaluation methods [16].

CONTRIBUTIONS

Eighteen quality contributions have been accepted in the workshop. They are categorized into three main groups:

(1) Case studies on work-oriented applications
- Wardlaw, Cox and Haklay on health care systems
- Gasik and Lamas on meeting room booking services
- Iknonen and colleagues on wellbeing and recovery management
- Hvannberg on accessibility and crisis management systems
- Sikorsik on the customer relationship management system for a call centre
- De Guzman on employee profile management

(2) Case studies on leisure-oriented applications
- Sutcliffe and Hart on art galleries
- Derboven on multi-touch interaction table
- Lárusdóttir on multiplayer online games
- Arhippaninen et al on 3D virtual music
- Johnson on virtual hotel games
- Väätäjä on city life exploration

(3) Methodological /theoretical frameworks
- Rantavuo and Roto on applying heuristic evaluation to study user experience
- Vermeeren and Cockton on analysing key concepts of the diffusion of innovation framework
- Cockton on the diffusion of the novel method - worth maps
- Bevan on usability maturity assessment and process improvement
- Jokela on the practical value of interviews
- Springett and Law on the possible integration of appraisal theory and action cycle

In-depth discussions in the workshop can shed light on these aspects with regard to the transferability of HCI methods. Future research challenges along this inquiry will be identified.

REFERENCES

[1] Andrews, M. Squire, C., & Tamboukou, M. (Eds.) (2008). *Doing narrative research*. Sage.

[2] Bygstad, B.,Ghinea, G., & Brevik,E. (2008). Software development methods and usability: Perspectives from a survey in the software industry in Norway. *Interacting with Computers*, *20*, 375–385.

[3] Hartson, R., & Pyla, P. (2012). *The UX Book*. Morgan Kaufmann.

[4] Jeffries, R., Miller, J., Wharton, C., & Uyeda, K. (1991). User interface evaluation in the real world: A comparison of four techniques. *Proc. CHI'91*, 119-24.

[5] John, B. E. (2000). Learning and using the cognitive walkthrough method: a case study approach. *In Proc. CHI'95* (pp. 429-436).

[6] Følstad, A., Law, E. L.-C. & Hornbæk, K. (2010). Analysis in usability evaluation: An exploratory study. In *Proc. NordiCHI 2012*, 647-650.

[7] Furniss, D. (2008). *Beyond problem identification: Valuing methods in a 'system of usability practice'*. PhD Thesis,: University College London.

[8] Gulliksen, J., Boivie, I., & Göransson, B. (2006). Usability professionals: current practices and future development. *Interacting with Computers*, 203-261.

[9] Lárusdóttir, M.K., Cajander, A., & Gulliksen, J. (2012). The big picture of UX is missing in Scrum projects. In *Proc. I-UxSED 2012*.

[10] Nonaka, I., & Krogh, von G. (2009). Perspective—tacit knowledge and knowledge conversion. *Organizational Science, 20*, 635-652

[11] Schön, D. A. (1984). *The reflective practitioner: How professionals think in action*. Basic Books.

[12] Suchman, L. (1987). *Plans and situated actions*. Cambridge University Press.

[13] Tullis, T., & Albert, B. (2008). *Measuring the user experience: collecting, analyzing, and presenting usability metrics*. Morgan Kaufmann.

[14] Venturi, G.,Troost, J.& Jokela, T. (2006). People, organizations, and processes: An inquiry into the adoption of user-centered design in industry. *Int J, HCI21*(2), 219-38.

[15] Wixon, D.R. (2003). Evaluating usability methods: why the current literature fails the practitioner. *Interactions 10*(4), 28-34.

[16] Woolrych, A., Hornbæk, K., Frøkjær, E., & Cockton, G. (2011). Ingredients and meals rather than recipes: A proposal for research that does not treat usability evaluation methods as indivisible wholes. *Int. J. HCI* 27(10), 940-70.

Adaptation of Method-resources Between Projects: A Case Study From a Dynamic and Complex Work Domain

Jessica Wardlaw
University College London
Gower Street, London, WC1E 6BT, UK
j.wardlaw@ucl.ac.uk

Anna L Cox
University College London
Gower Street, London, WC1E 6BT, UK
anna.cox@ucl.ac.uk

Muki Haklay
University College London
Gower Street, London, WC1E 6BT, UK
m.haklay@ucl.ac.uk

ABSTRACT
In this case study we describe how method-resources were reconfigured across three design and evaluation projects conducted by an in-house design team within the same company during a six-year action research collaboration with academics from the field of Human-Computer Interaction (HCI). This case study specifically focuses on the reconfigurations that occurred in participant recruitment, task selection, reporting format and problem identification between the three projects. The underlying contextual factors behind the reconfigurations, in particular the application domain, organisational factors and project constraints, will be discussed to give unique insights into the realities of design work from within a single organisation over the six-year collaboration. This case study demonstrates the complexity of comparing methods across projects, particularly within dynamic and complex work domains, and that existing attempts may be too simplistic because they fail to account for these factors.

Author Keywords
Action research; design and evaluation methods; resources.

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

General Terms
Design; Human Factors.

'OLD' METHOD-RESOURCES IN NEW CONTEXTS

In this case study we describe how the same method-resources were applied differently across three projects conducted by an in-house design team within the same company during an action research (AR) collaboration. AR encompasses methods and approaches for collaborative research with partners towards addressing problems they experience; this is done through cycles of planning, action and reflection, which offer HCI the opportunity to address gaps between theory and practice [4]. It is interesting to compare the projects since the business environment significantly changed during the collaboration, imposing very different constraints on the projects despite their strategic importance to the company. The resources that this case study will focus on are: participant recruitment (finding the right type and number of participants), task selection (specifying tasks for inspection or user testing), reporting format (communicating problems and solutions for subsequent analysis, evaluation auditing, iteration and customer communication) and problem identification (tools and approaches for identifying/discovering problems) [16].

Dr Foster Intelligence (DFI) is a public-private partnership in the United Kingdom (UK) health informatics sector that provides independent health and social care information to healthcare managers and clinicians for the improvement of clinical effectiveness and efficiency. DFI was formed in 2006 as a partnership between the National Health Service (NHS) Information Centre and Dr Foster Ltd. DFI has produced a range of web-based data analysis tools which give NHS managers access to the Hospital Episodes Statistics (HES) database that contains admitted patient care data from 1989 onwards and outpatient attendance data from 2003 onwards. Whilst live access to a database of 825 million hospital records presents many challenges, particularly with users that vary greatly in requirements and geographic location, it also presents great opportunities that are unavailable in any other country's health system.

In 2010, however, after a change in Government, the Department of Health bought the NHS Information Centre's shareholding and announced a strategic review of the future of DFI [9]. Following this review, and the Government's Spending Review, urging Departments to maximise value from assets that do not need to be held in the public sector, it was announced that DFI would be marketed for sale [11]. This sale is still being negotiated.

The company therefore faces many challenges to maintain their position as a leading provider of health informatics in the UK. Since the company began, the market has become more competitive and many trusts will develop internal solutions to save money. Financial constraints have resulted in customers having more complex and changing needs and demanding more choice. The usability of health informatics tools, and how well they meet users' requirements, is thus an increasingly important factor when health organisations are deciding whether or not they will invest in them.

Project 1: Obtaining user requirements for and evaluating Population Health Manager

Primary healthcare services in England (e.g. doctors, dentists, opticians and pharmacists) are managed by local Primary Care Trusts (PCTs). PCTs control 80% of the NHS

budget [2] to determine and provide the health services local communities need, including hospitals. Many PCTs use DFI's Population Health Manager (PHM) tool for this work, which provides PCTs with the information to:

- Understand the local population and develop segmentation models of their health needs
- Identify and analyze local health inequalities to target unmet needs or gaps in care
- Monitor admission trends, forecast population health needs and predict future health trends.

PHM offers various datasets, which are regularly updated, along with the facility for users to upload their own local datasets. Maps of PHM data can identify spatial inequalities in the provision of health outcomes to inform the location of appropriate services and interventions, in addition to understanding patient referral patterns (Figure 1).

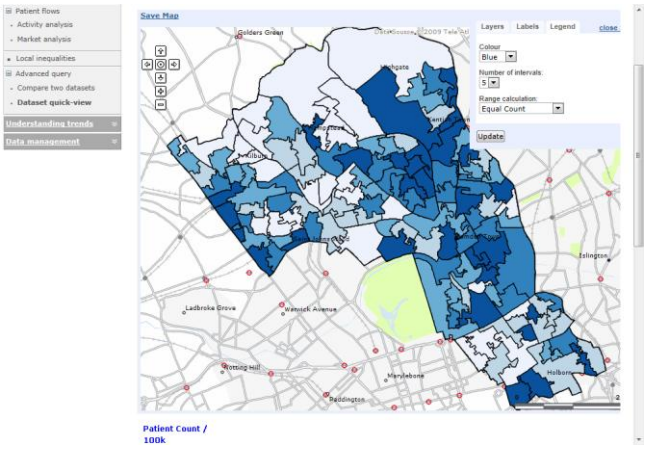


Figure 1. The mapping interface for PHM

During the PHM project, an online survey was used to establish potential users' preferences for the cartographic presentation of the data and thus inform the design of a new mapping interface. Questions presented two or three maps covering various cartographic aspects of the maps including data classification, number of ranges, colour schemes, the representation of point data, raster or vector data for the background map and mapping multiple datasets. For each aspect participants were asked to answer a question about the data that required interpretation of the map and to identify which map enabled them to answer it more easily and which option they preferred. Survey results were then incorporated into the software design. The final interface was then inspected using Heuristic Evaluation [5] and Cognitive Walkthrough [8].

Project 2: Developing company Personas

Following the PHM project a need was recognised to improve the developers' understanding of the end users of DFI products, since they did not have regular opportunities to meet the end users to understand why the tools were used and the development team personnel was often changing. To achieve this, a specific project was initiated to create personas of the key DFI users in which a variety of methods were combined according to the resources that were available. Fourteen semi-structured interviews with key DFI users were supplemented with information from database server log files (reflecting usage of the entire user population) to confirm which job titles represented the most frequent users and which parts of the tools were used the most. In addition, a user-generated screenshot survey required users to take a screenshot of their entire computer screen whilst using a DFI tool which revealed rich information on users' working environments and taskflow.

These methods were specifically chosen due to the wide geographical distribution of users and the relevance and richness of the information they could provide with limited resources. These personas were used, in part, to inform the redesign of DFI's flagship product, which was to be relaunched under the name Quality Investigator (QI). For this product user testing was also carried out.

Project 3: User Testing of Quality Investigator (QI)

QI is a web-based tool that monitors quality outcomes and patient safety by assessing clinical, process and coding factors. This was developed three years after PHM. Its user interface comprises tabs for Mortality, Length of Stay and Readmissions, all key indicators of clinical quality and efficiency (Figure 2). A dashboard highlights a hospital's 'CUSUM alerts' for diagnosis and procedure groups; negative CUSUM alerts (indicated by red bells) are given when indicators diverge sufficiently from expectations to suggest a systematic problem. 'Relative Risk' also provides the observed cases as a percentage of the risk-adjusted expected (reflecting case mix and national average). This permits analysis of patients by diagnosis or procedure group and comparison of clinical performance. The five diagnosis and procedure groups with the highest 'observed' (number of cases within the selected dataset) exceeding 'expected' (expected cases given the case mix) and crude rate (observed cases as a percentage of volume) are also shown.

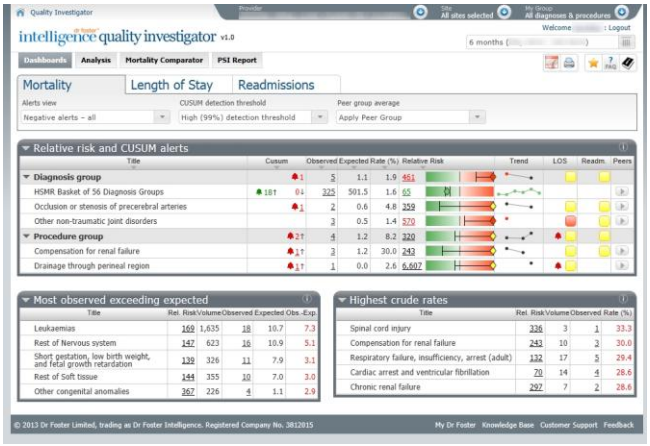


Figure 2. Quality Investigator dashboard

Before these studies' resources are compared, Table 1 presents the acronyms introduced thus far for reference.

Acronym	Full Term
DFI	Dr Foster Intelligence
PHM	Population Health Manager
QI	Quality Investigator

Table 1. Acronyms.

Stories of Transfer: Triumph or Tragedy or Both?

In this paper we examine the reusable HCI method-resources across the projects described to identify where they have proven to remain useful and compatible or require substantial modification when applied across the usage contexts of the projects.

Participant recruitment:

Each of the projects required recruitment of participants.

At the time of the PHM project, there were very few resources for usability work and usability work had little status within the organisation. Furthermore, it was a new product in a new market for DFI, so there was only a limited network from which to recruit participants. Survey participants were therefore recruited by emailing users that were suggested to the researcher by colleagues.

Recruitment of key users for interviews to inform the personas was negotiated and managed with the Customer Service Managers, who regularly meet with users, so that they could approach any potential participants initially.

In contrast, participant recruitment for the QI project was assisted. This was in recognition, by the organisation, that users must be engaged in the design process because contracts were at risk of non-renewal.

It can be concluded that the facilitation and effectiveness of participant recruitment has improved across the three projects. Whilst Amazon vouchers were offered for participation in the maps and user-generated screenshot surveys, and the suitability of some interviewees for the personas can be questioned because of their very limited experience of the tools, participants in the user testing were motivated to take part by the opportunity to be involved in the development and direction of the new tool and to try it ahead of its launch. The low number of users who tested QI, however, may have impacted upon the reliability of the results given the complexity and breadth of users of the application domain [15].

Task selection:

Two of the projects required the creation of tasks to evaluate the tools being developed. Data collected to create personas also included identification of real users' tasks.

Tasks were not well defined for the PHM project because it was a new product in a largely untested market for DFI. In order to design a task for the Cognitive Walkthrough it was necessary to look at job descriptions of the target audience, in an attempt to understand the type of work that they did. A DFI colleague who supported a particular local health organisation to carry out tasks similar to those PHM was designed to support was also available to consult on the types of reports that his client wrote.

To create personas, we used Contextual Inquiry [14] consisting of semi-structured interviews and unstructured observation sessions during which participants demonstrated a task that they commonly performed using a DFI tool. This produced a range of closed and open tasks so arguably provided a much more realistic picture of the tasks users aim to accomplish with the tools.

For the user testing of QI, tasks were designed based on information provided by customers as to the tasks they would like to be able to perform with the new tool, which included new functionality, to which the project team gave priority and improvements on existing functionality. Some tasks could not be completed during the initial user tests because of the development stage of the tool; however these tasks were implemented for subsequent tests in addition to some quick fixes to issues that were identified during the initial tests. There was also a second phase of user testing in order to test a more completed version of the tool.

The ability to select realistic and appropriate tasks to carry out the methods has gone from tragedy to triumph during the three projects through a growing understanding of the application domain and users' requirements. However, there has been no opportunity to reuse the tasks generated as each project has focused on a different tool that supports a different part of the users' work.

Problem identification:

Each project resulted in the identification of numerous usability problems, however they varied dramatically in terms of their success with this. The Heuristic Evaluation of PHM exploited a structured report of 296 heuristics available online [5] loosely grouped according to Nielsen's ten usability heuristics [7]. Each usability problem identified was then assigned a severity rating according to the classification in [10]. Similarly the Cognitive Walkthrough followed the format outlined by [8], with the addition of a fifth evaluation question that asked what the system provided beyond the normal method by which users would carry out the task. These methods together identified 32 usability problems; 12 rated as irritants, ten rated as moderate, nine rated severe and one as unusable.

To create personas, interviews were conducted in which users revealed their frustrations with existing tools. Additionally, participants were asked to demonstrate a typical task they perform on the website, which revealed some additional problems; this was recorded using video

capture software. Although activities conducted to create personas are not necessarily designed to elicit usability problems with existing software, this was a serendipitous outcome of this project. This demonstrates extension of the textbook scope of Contextual Inquiry and reproduces some of the view developed by [6].

The QI methods included a user study, which identified a large number of problems. It was a relatively easy job to prioritise the problems: for example, feedback on results was requested for a Monday morning scrum meeting and results were quickly compiled after Friday's final user test. Standard usability evaluation methods are known to vary in terms of the number and severity of problems identified. The expertise of the evaluator is also known to influence these outcomes. It is likely that all of these factors impacted our projects. An unexpected triumph was that interviews conducted for the purpose of persona generation also revealed usability problems that could be addressed by DFI.

Reporting format:

The reporting format also varied between the projects.

The personas were first compiled into a PDF file comprising of a page for each persona that included details of their goals, working environment, typical behaviour, attitudes and skills, in addition to a photo and some personal information to bring the personas to life. These were then printed as A3 posters for the walls of the office. Subsequently a more detailed report of the main themes that arose in the interviews was written to extend the scope of information conveyed by the personas; this comprised of a summary of tool usage patterns, who the users are (their job titles, roles and responsibilities and main motivations for the tools) in addition to problems highlighted with the data in the interviews (transparency of data source, data quality, timeliness of data, unclassified data, data complexity, analysis and information presentation) and recommendations that interviewees gave for improving the tool functionality. The results were communicated to the rest of the team through a presentation of the personas and summary of the report (with a focus on the issues that interviewees identified with using the tools) at the Product Development team monthly meeting; both the personas and report were shared with the team after this meeting as well as the interview transcripts, so that developers could develop an understanding of the language used by the users.

In the new business context of QI, since the user testing was formally part of the development process a much more concise report was provided. In addition to a written report, highlights of the user testing sessions were communicated at a project meeting through a presentation that included a summary video. This video was designed to show examples of both unsuccessful and successful task completion so as not to discourage the developers.

The reporting format has largely been configured according to its audience and therefore been successful in the transfer of redesign proposals through to development. For PHM the audience was primarily the designers working with the developers, for the personas the Product Development team, and the results of the user testing were communicated to the project leads before being prioritised and put through to development at the weekly scrums. However, there was no opportunity to reuse existing formats for reporting. Development of a standard reporting format to be used within DFI would facilitate reporting in future projects.

ANALYSIS OF CONTEXTUAL FACTORS

Over the time course of the three projects a number of key contextual factors changed. We now discuss their influence on method-resources and on the success of applying the design and evaluation methods in practice. To do this we use the classes of resources defined in [16].

Axiological resource types

Axiological resources refer to the values that motivate an approach, for example clients' needs and expectations from a method and corporate culture and values [16]. Across the project presented, the action research approach taken necessitates consideration of the clients of DFI (i.e. the end users of DFI's systems) and the client within the collaboration (i.e. DFI); both relationships require careful management of expectations at the individual level and clients will place different values on the methods [3]. This was evidenced in the participant recruitment and test protocol for the user testing of QI; DFI were keen not to give the participants the impression that any suggestions they made for improvement would be implemented. In addition, the nature of the collaboration requires that there is a mutual understanding that any one method would not provide the 'silver bullet' that DFI might hope for. The reporting formats were also sensitive to how long each project allowed for analysis and reporting of results. The end users' needs and expectations for PHM were not clear since there was no formal requirements gathering process; moreover, the organisation's expectations of the final mapping interface had to be managed according to what functionality could be implemented given limitations arising from the underlying architecture of the product.

At the team level, the development team for QI was much smaller than the PHM team, which has facilitated more effective and efficient communication of evaluation results. More fundamentally, changes in development team personnel resulted in a change in developers' skills between the two projects. Agile developers were recruited for the QI project, which meant that they were much more accustomed to an iterative design process. As has been noted, a user experience expert was also recruited for the QI project who brought with them their own knowledge and expertise in designing and running user testing sessions.

At the organisational level, there has been a change in the perception to user testing towards employing it as a tool to build and maintain client relationships. Client relationships

are largely protected, as would be expected for any commercial development organisation with busy users with whom they often have to consult due to the complexity of the domain; this impacted upon participant recruitment particularly during the first stages of the collaboration. This was especially important since persona interviews were pitched as an exercise for the Engineering Doctorate that forms the basis for the collaboration, despite the benefits for DFI being explained in full before users consented to their participation. Participants for user testing were recruited from users whose contracts were about to end and the organisation had reason to be concerned that they might not renew. The aim was to encourage these participants to renew by exposing them to upcoming developments. This in turn has increased management support for usability work; the visions and values of key stakeholders can be an important influence on how other resources are assembled and configured in design work [13]. This approach to participant recruitment is partly necessary due to the complexity of the work domain but in stark contrast to the approach of many design consultancies that are able to send screening questionnaires to many potential participants.

It is important to note that whilst there was not a shortage of budget for the PHM project overall, due to lack of awareness around usability, usability work itself was not allocated a separate budget. In this case study more project resources were allocated to usability when the company was performing less well financially than when it was performing well financially. The need for improvement in the quality of design work largely stemmed from the increased market pressure described, and for improvements to be effected there had to be a change in the status of usability work across the organisation and recognition of its importance, particularly from senior management. This resulted in the recruitment of knowledge and expertise in usability, integration of design and evaluation methods into the development process and more successful use of resources and methods.

Expressive resource types

Expressive resources are those that communicate evaluation findings [16]. In this case study, the format and medium for reporting the results of design and evaluation methods required adaptation with the shift from a Waterfall development approach to an Agile approach, which demands a faster and more concise reporting format. This demonstrates that design and evaluation methods can be more effective in dynamic contexts if they support the rapid analysis and feedback of results. As Sy (2007) describes, for the Agile development process results were reported through the weekly scrums, whereas the Waterfall development approach enabled the writing of much more detailed reports [12]; the usability process was much more informal for PHM compared to how it was integrated within the development process for QI.

Knowledge resource types

Knowledge resources refer to knowledge of the system under evaluation, users and their abilities and tasks, and the application domain [16]. Such resources can impact upon all other method-resources. Growth in knowledge resources during the collaboration has directly resulted in more successful transfer of resources between projects.

Over the course of the collaboration the primary researcher developed her own experience of using design and evaluation methods, through guidance from academic experts, Masters courses and various workshops; this will have unavoidably coloured the way in which the methods have been applied and the projects have been planned. For example, the list of heuristics used and the format of the Cognitive Walkthrough were deliberately selected to be highly structured to provide additional support for problem identification. One particular consequence of this was that the heuristics used were partly adequate but also partly inadequate for the complexity of the interface inspected; many heuristics were assigned ‘not applicable’ and therefore the support provided by the heuristics used for problem identification in this case is questionable.

It should also be noted that between the PHM and QI projects three years had passed, during which a user experience expert was recruited by DFI who brought his own expertise in managing client relations when conducting user testing. His experience was particularly useful in the design of the testing protocol, for example letting the clients talk about their general experience of using the original DFI tool before introducing them to the new design. This was in addition to the growth in knowledge resources within the product development team through training, experience and arrival of new staff.

Finally, for information systems such as the ones described, knowledge of how the users think about the data is manifested in the database and software architecture. In this case study, easy modification of the user interface was found to depend on this knowledge much more than its separation from the software architecture [1]. At the time of PHM the architecture of the underlying database made it fundamentally very difficult to implement some redesign suggestions but in the three years between PHM and QI the architecture was reconfigured with an explicit aim of making changes much easier and more stable to implement. This can be attributed to the complexity of the application domain and had a big impact on the resulting design decisions made and whether redesign suggestions were put through to development. This improved flexibility has furthered receptiveness of usability work within DFI.

CONCLUSION

Methods and resources were employed across design and evaluation projects: a survey to inform the design of a mapping interface and its subsequent inspection using Heuristic Evaluation and Cognitive Walkthrough; the

creation of company personas using database server log files, a screenshot survey and interviews; and finally the user testing of an interface whose redesign was informed by these personas. Over this period, a number of significant changes took place within and outside of the organisation. The influence of these changes was that participant recruitment, problem identification, the reporting and dissemination of results and task selection have had to be reconfigured according to local resources over the course of the collaboration, with increasing success as reported.

The reality of design work illustrated by the changes in business environment described in this case study is that designers consider the "ingredients" available to them before deciding which "recipe" to follow [16]. Recipes can come from their own recipe book, or the "HCI" recipe book. The ingredients, and importantly the cooking methods/utensils, available are determined by the application domain, organisational factors and project constraints. This case study demonstrates that local resources can be more influential than those indicated by the textbook versions of methods and more important than any financial cost of the ingredients and cooking utensils: development context (especially the process being used and how methods fit within this), organisational culture (supported from the highest levels of the organisation), knowledge resources (the expertise currently available) and the clients' expectations and needs were especially important. Studies that compare methods used in different contexts frequently ignore such factors, which case studies such as this suggest is to their detriment.

We conclude that this action research project has been of great benefit to DFI in raising the awareness and status of usability at the organisation and integrating methods into the design and development process; this has included making the method-resources available for the high quality usability work required in this complex domain.

ACKNOWLEDGEMENTS

This work was undertaken for an Engineering Doctorate supported by the Engineering and Physical Sciences Research Council (EPSRC) and DFI, in addition to an Industrial Fellowship from the Royal Commission for the Exhibition of 1851. We further thank all participants in the studies described and the three reviewers for their feedback.

REFERENCES

[1] Bass, L. and John, B.E. 2003. Linking usability to software architecture patterns through general scenarios. *Journal of Systems and Software*, 66(3), 187–197.

[2] Department of Health 2006. *Health reform in England: update and commissioning framework.*

[3] Furniss, D. 2008. *Beyond Problem Identification: Valuing methods in a "system of usability practice"*. University College London.

[4] Hayes, G.R. 2011. The relationship of action research to human-computer interaction. *ACM Transactions on Computer-Human Interaction*, 18(3), 1–20.

[5] Heuristic Evaluation – A System Checklist: 1995. *http://www.stcsig.org/usability/topics/articles/he-checklist.html*. Accessed: 2013-01-18.

[6] McDonald, S., Monahan, K. and Cockton, G. 2006. Modified contextual design as a field evaluation method. *Proceedings of the 4th Nordic conference on Human-computer interaction changing roles - NordiCHI '06* (New York, USA, Oct. 2006), 437–440.

[7] Nielsen, J. 1994. Heuristic Evaluation. *Usability Inspection Methods*. J. Nielsen and R.L. Mack, eds. John Wiley & Sons, Inc. 25–62.

[8] Polson, P., Lewis, C., Rieman, J. and Wharton, C. 1992. Cognitive walkthroughs: a method for theory-based evaluation of user interfaces. *International Journal of Man-Machine Studies*, 36(5), 741–773.

[9] Review of future of Dr Foster Intelligence: 2010. *http://www.dh.gov.uk/en/MediaCentre/Pressreleases/DH_118452*. Accessed: 2012-04-24.

[10] Rubin, J. and Chisnell, D. 2008. *Handbook of usability testing: how to plan, design, and conduct effective tests*. Wiley Pub.

[11] Sale of Dr Foster Intelligence: 2010. *http://www.dh.gov.uk/en/MediaCentre/Pressreleases/DH_121537*. Accessed: 2012-04-24.

[12] Sy, D. 2007. Adapting Usability Investigations for Agile User-centered Design. *Journal of Usability Studies*. 2, 3 (2007), 112–132.

[13] Uldall-Espersen, T. 2008. Visions in software development: Achieving value in organizations. *Proceedings of CHI '08 workshop: Values, value and worth* (Florence, Italy, 2008).

[14] Wixon, D., Holtzblatt, K. and Knox, S. 1990. Contextual design: an emergent view of system design. *Proceedings of the SIGCHI conference on Human factors in computing systems Empowering people* (1990), 329–336.

[15] Woolrych, A. and Cockton, G. 2001. Why and When Five Test Users Aren't Enough. *Proceedings of IHM-HCI 2001 Conference (Volume 2)* (Toulouse, France, 2001), 105–108.

[16] 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. *International Journal of Human-Computer Interaction*, 27(10), 940–970.

Lean Design for Good User Experience

Valeria Gasik
Tallinn University, Narva mnt 29
10120 Tallinn, Estonia
valeria@sokeri.org

David Lamas
Tallinn University, Narva mnt 29
10120 Tallinn, Estonia
david.lamas@tlu.ee

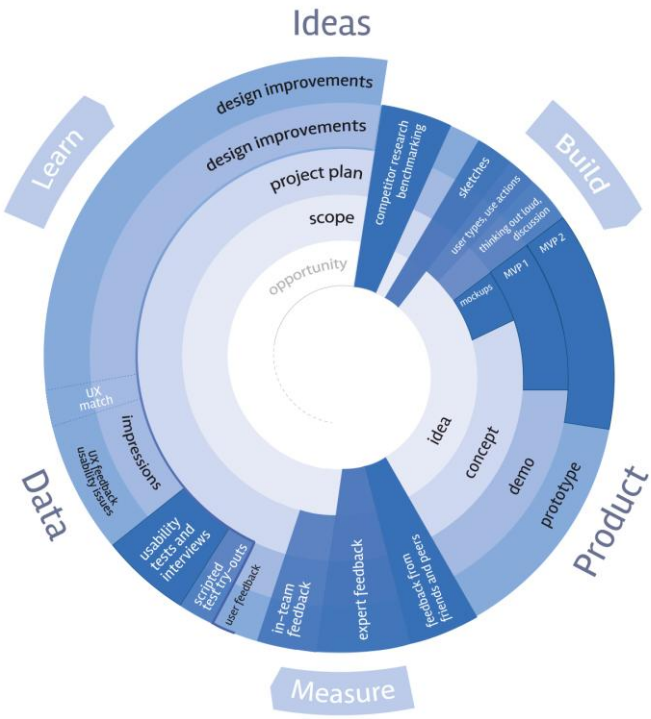


Figure 1. The Lean spiral [3].

ABSTRACT

In the saturated market of online commerce, success of a new service is tightly connected to the quality of user experience. A company cannot design positive user experience as such. Instead, one can design for certain key factors that are related to the typical usage of the service. This design process is a business challenge, as it has to be balanced with organization's own values, goals and resources.

Author Keywords
Lean; user experience; transfer

In the work herein described, this challenge was framed with two specific research problems: the first question was in approaches that could support designing for good user experience in an early stage project; the second question was whether and how Lean principles could guide this design process. Results support the proposition that applying Lean principles for designing online services facilitates the achievement of good user experience. As an outcome, the outlining of an overall framework for using Lean principles in the implementation of similar projects is proposed.

ACM Classification Keywords
H.3.5 Online Information Services; Commercial Services/Web-based services

H.5.2 Information interfaces and presentations (e.g., HCI); User Interfaces; User-centered design

General Terms
Lean, User Experience

INTRODUCTION

This work explores the idea of achievement of a *good user experience* in the context of the design and development of **Roomforit.com**, a localized online service concept for meeting rooms booking. The described research was started with two assumptions:

1. When designing for good user experience, even a small team with limited resources could create a valuable online service concept; and

2. Design process of such potentially valuable concept could benefit from Lean principles.

Based on these assumptions, the following research problems were formulated:

1. What approaches could be used to support designing for good user experience in the context of the selected project?

2. How can Lean principles guide design process of this project?

USER EXPERIENCE

Based on the literature review [e.g. 12, 15] several approaches could have been selected to support the design and development for positive UX as well as assessment of the perceived UX of **Roomforit.com** concept. For instance Roto et al. [15] have collected over 80 methods for designing for *User Experience (UX)*, which were categorised, among others, by type, development phase, information provider and the length of period when user experience is studied. The selection of right approaches depends on the level of decision-making, scope of interest and time frame of the reflection. It is clear that all aspects of a service could not have been applied in the framework of the project herein reported; it is even arguable whether one should do so in any other practical case. Rohner [12] noted that there is indeed no point to use all possible methods — rather one should select methods based on the questions they are aimed to answer. What could be withdrawn from Roto, Law, Vermeeren & Hoonhout [14], Jetter and Gerken [8], Hassenzahl [6] and others is that there is no magic trick for designing an ultimately good user experience but there are some underlying principles that could guide the design process. It is likely that a service can not provide everything to everyone but it might provide good settings to support most important user experience factors, e.g. Roto et al. [15] summarised: ”It is usual that a design team will only be able to deal with a few critical UX factors that influence the suitability of the design for a typical usage situation.”

In the light of the new online service concept **Roomforit.com**, outlining these critical UX factors, designing for contemplated, crucial UX facets as well as finding and applying relevant assessment approaches was part of the overall design process. This process was balanced with project’s business interests such as operating with a scarce budget and outlining suitable revenue models. As such goals are seldom regarded as main interaction design objectives, business aspects were studied from Lean principles’ perspective.

LEAN

Designing with Lean principles refers to searching ways to provide great customer value with efficiency but without compromising product quality. Lean was a namesake given in US in the late 1980’s to the concept of *Toyota Production System* (TPS) — a set of principles, which were iteratively shaped over a course of several decades.

According to the literature [e.g. 10, 17] TPS have been interpreted both as guiding principles (e.g. poka-yoke, kaizen) and a set of practical approaches (e.g. Kanban, 5S), aimed to solve the ancient question of creating value with efficiency. Lean, originating from car manufacturing, has often been systematised and occasionally codified to serve better the needs of other fields. Through time, Lean principles has been picked up and tuned by various industries including service, software development and most recently — project management and entrepreneurship oriented Lean startup. The main principle of Lean, derived from TPS, has not changed over time, although new or renovated approaches have been suggested, i.e. Kanban approach in Lean software and developing for *Minimum Viable Product* (MVP) in Lean startup. Every transformation faced praise and critique — at best Lean has been seen as a buttress for experimental hypotheses and validated decision-making, at worst as an expensive consultancy scam. To avoid this, Lean has been suggested to be viewed as a starting point for organisations that will develop and obtain their own principles [e.g. 2, 10].

CONNECTING LEAN WITH USER EXPERIENCE

Based on literature review, several connecting points with Lean ideology and User Experience might be found. Despite the differences in objectives of interaction design (i.e. good usability, positive experience) and business principles (i.e. increasing revenue), both approaches put user — or client — in the focus of the design. Nielsen, Norman and Tognazzini [11] proposed that user experience is”all aspects of the end-user’s interaction with the company, its services, and its products”, pointing out that company needs to make numerous assumption of how its service is perceived by the user. For instance through interviews and usability testing it might be discovered that users understand service’s concept and are able to complete main tasks well, but in real life they would not use the service because in their opinion, it is not *cool* and attractive enough [8]. As there are enormous amounts of possible individual and dynamic assessment attributes, a guiding framework, which will create a focus for design project, is in place. When the goal of the project is to find optimal points of providing good user experience and building a valuable business concept, Lean can be seen as another supportive set of values.

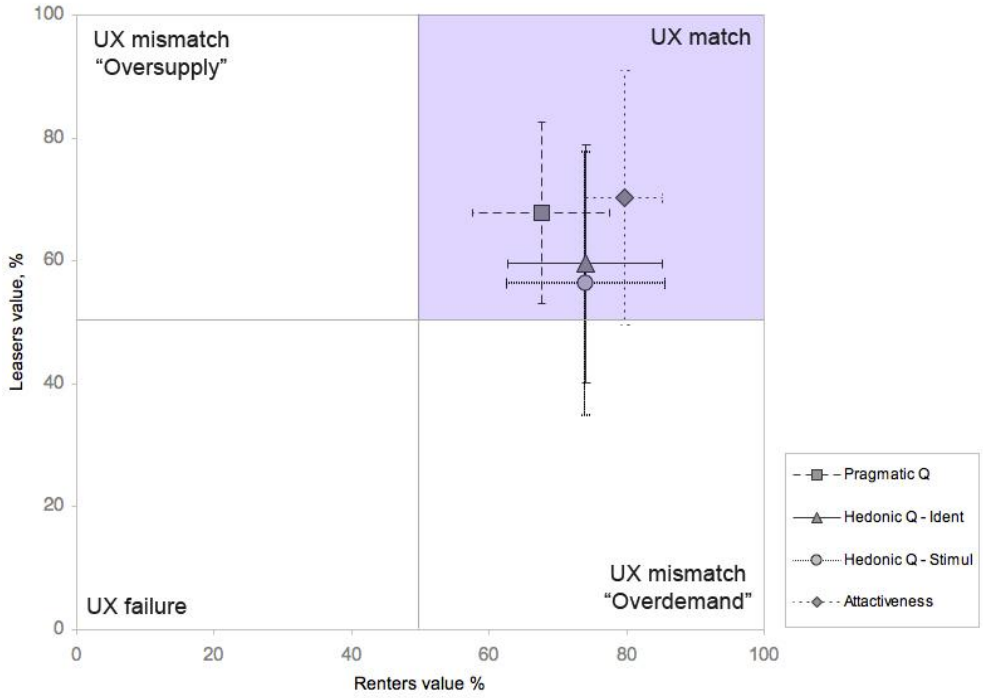


Figure 2: Comparison of perceived service quality values by renters and leasers. Data is based on two AttrakDiff [1] word pair Single Evaluation survey results. Error bars indicate standard deviations within 4 word pair categories (7 word pairs in each). UX match (colored top right square) means that both sellers and buyers would likely to use the service in terms of all analyzed quality and attractiveness criteria. [3]

Designing for good user experience and designing with Lean ideology both require experimental iteration and as Ries [12] noted: "in-person customer observation". It could be argued — though validation of such claim lacks scientific support — that Lean can be beneficial as a guiding principle when designing for good user experience. Lean principles of designing out waste and focusing on customer's value can help to maintain a balance when team is faced with various choices and need to decide whether to proceed or pivot. Lean startup's additional practical suggestions of e.g. building a Minimum Viable Product and striving to achieve validated learning can administer to form a project framework and set overall goals already at the stage of concept creation.

THE PROJECT

Roomforit.com is an open marketplace for listing and renting meeting rooms. This online service concept was introduced in the design research through inspection of four phases — idea, concept, demo and prototype. Reflection focus was set on the selected UX approaches and Lean principles, i.e. attention to customer value, recognizing and redesigning out *waste* as well as validated learning in terms of building MVPs and evaluating them with relevant UX assessment methods. The process is inspected in the Lean spiral, Figure 1. Figure is based on iterative Build-Measure-Learn cycles [3].

The first phase describes devising of the strategic business idea. Benchmarking, competitor research and sketching, among others, supported formulation of business proposals that were reflected with feedback from experts and friends. Re-assessment through the lens of Lean supported finding a feasible direction for the project, i.e. seizing limitation based on available resources as well as focusing on the most relevant needs of users in the light of this potential business.

The concept phase covered more detailed examination of the business idea, supported with various interaction design approaches, e.g. outlining user types and composing visual mockups as well as collecting more feedback. The focus set with Lean was acknowledged in derived strategical and pivotal changes, such as leaving hot desk renting out from concept's MVP, thus concentrating only on UX of renters and providers of meeting rooms.

Paper sketches, wireframes, *role playing* use scenarios and other approaches accompanied with collected feedback were employed in the third phase of designing a live demo, published for selected audience — the first MVP. For a holistic assessment of user experiences by both renters and users, second MVP with relevant mockups was created during the prototype phase.

Business insights gathered during concept and design phases led to strategical discoveries as well as systematic

recognition and reduction of *waste*. For instance during prototype phase, service's characteristic was tuned — **Roomforit.com** was repositioned from a mediating *middle-man* service to a platform that supports free communication between two main user groups. This change drastically the amount of design work needed for building second MVP mockups.

The critical UX factors of **Roomforit.com** concept were crystallised during demo and prototype phases. It was concluded that success greatly depends on a match of perceived good user experience by both groups, renters and room providers (leasers). Specific critical attributes, e.g. usable, credible and friendly, which were outlined with consistent benchmarking and feedback from peers and colleagues as well as practical design, were helpful for design orientation but trivial in the light of evaluating the overall success rate of the intended **Roomforit.com** concept values. Based on this understanding, UX assessment was done with six representatives, three from both group — renter's and leaser's. All individual meetings included scripted usability tests, targeted to assess typical usage tasks and AttrakDiff [1] Single Evaluation word-pair surveys, aimed at measuring attractiveness of the service in terms of usability and appearance. As AttrakDiff [1] visualisations did not clearly expressed the overlap and clustering of these two survey results, for more holistic analysis, data points were inspected in UX Match Matrix visualisation, described in the Figure 2 [3].

REFLECTION

As mentioned in the introduction, we focused our research questions on designing for good user experiences using Lean principles.

While usability tests and survey results showed evidence regarding overall positive UX, improvements and new evaluations have to be made in the future. For instance usability tests pointed out that there were no critical usability flaws, although participants were struggling with some of the tasks. Also some dispersion in AttrakDiff [1] word pair survey results was witnessed, i.e. leasers' answers were less unified than those of renters'. In addition, there was some discomfiture in the survey semantics: for example results for words *undemanding* and *challenging* were dubious. Nevertheless it was concluded that selected combination of UX approaches together with Lean principles and Lean startup approaches such as designing for MVP, has supported the positive outcome. Because both test groups had generally positive experience, it might be concluded that **Roomforit.com** concept was on the right track. This result leads to critical discussion about benefits of selected approaches as well as the role of Lean.

The Lean spiral in Figure 1 [3] highlights frequently used approaches, most essential evaluation methods as well as products, data and ideas generated throughout design process of **Roomforit.com**. Graph is based on Lean startup model *Build-Measure-Learn*. The original figure is extended from a circle to a spiral form, which represents in more details the iteration process of project's main four phases. In the **Roomforit.com** project, most useful and used internal approaches to were benchmarking and competitor observation, sketching, low and medium-fidelity prototyping, written or orally communicated user actions and *thinking out loud about design* i.e. walking though use scenarios while *role-playing* a renter or a leaser. In-team feedback is presented in both, Build and Measure, sections: direct discussion was typically related to some details of design-in-progress, while feedback was closely connected to evaluation of design solutions that were already made.

External approaches that were done in collaboration with people outside the team included gathering feedback from friends, peers and domain experts. Secondary data of unstructured and contextual feedback was collected frequently throughout the whole process. Real user feedback was collected based on the live demo and clickable prototype mockups. Experimenting with the first two scripted tests gave confidence in conducting meetings with test participants; they also served well for collecting the first direct feedback from potential **Roomforit.com** users.

Although selected approaches worked fine in the context of this project, some limitations should be acknowledged. Standalone, most of the approaches are quite weak: for instance relying on experts' opinions might only give professional perspective but not reveal the everyday problems of regular users. It is viable to note that retrospective analysis based on private blog notes and memories might distort true impact of each approach at various stages. It should be also pointed that while Lean values were communicated to the team, during the idea and concept phases one person did most of the work of *UX researcher* [9] — this had a limiting impact on the efficiency of the process.

Some of the design means visualised in the Lean spiral (Figure 1) might be debatable. For instance *thinking out loud about design* and "role-playing" users were not mentioned in the literature and thus could be considered as not directly related to UX approaches or Lean principles. In terms of Lean, for instance Seddon and O'Donovan [16] pointed out that even if some Lean approaches are not listed they should not be considerate irrelevant. In the perspective of UX, e.g. Goodwin [5] proposed innovative UX evaluation methods during design process. Theatre and drama as part of design process has been also discussed by experts in the neighbouring field of service design [e.g. 4, 7]. During the project, talking out loud about design and discussing details with the team resulted in many insightful thoughts.

Lean startup principles strongly recommend rapid prototyping and reflecting the design with user feedback as early as possible. Rapid prototyping is familiar from Agile methods and interestingly, this recommendation seems to

be a new addition to the original TPS principles, which, in contrast, seem to prefer iterative perfection and quality over speed. Real user feedback was collected only on high fidelity mockups and live demo, not earlier. It was rationalised that secondary data was enough for the first concept and design phases. Meeting rooms booking service **Roomforit.com** resembles services for booking hotels, flights, movie tickets, gym facilities and so one — a process familiar to many. In this sense, a general assessment and worthy feedback could be given by many non-professionals.

Third rationalisation was rather paradoxical and likely most debatable. Gathered information indicated that some people working with booking systems were quite frustrated with them. Introducing and testing raw ideas and poor visual representation of a product that has already been out in various forms but which did not please its users, was seen as waste of resources and time for both potential customers and project team. Aim to establish and design out commonly known problems, before testing with real life users, was seen important.

Roto, Law, Vermeeren & Hoonhout [14] noted that user experience is dynamic and might change over time. What participants experience during usability tests or interviews might be different from what they would perceive when using the service in the context of their everyday life. From the point of view of **Roomforit.com** project this means that while user's reflections were positive at the moment of evaluation, new issues will continuously occur when service is published and used in a the real life context. Because of this, service should continue to validate various design and business hypotheses as it was done in the first round of concept creation. This type of relentless reflection and continuous improvement was presented by original Lean principles and adopted in Lean startup's model of *Build-Measure-Learn.* Such project values are thus likely beneficial also for the future development.

While this work does not suggest the adoption of precise sequence and combination of project's approaches in other online service conceptualization projects, principles and components presented in the spiral (Figure 1) could work as discussion mobilizers in similar works. The UX Match (Figure 2) concept could be beneficial for comparison of two or more important user groups, securing a better potential success of similar online services.

CLOSING REMARKS

As an overall result, it has been confirmed that applying Lean principles for design of such new online service allows achievement of good user experience. Integration of Lean with designing for good UX is possible and such overall concept could be extended on any relevant service development. The spiral (Figure 1) can support build-up and development of any such typical project. The UX Match concept (Figure 2) can support evaluation result comparison of two or more user groups.

ACKNOWLEDGMENTS

The work was supported by Aalto Center for Entrepreneurship (ACE) "Aalto Pre-Seed Funding".

REFERENCES

1. AttrakDiff. (2011). Retrieved from http://www.attrakdiff.de/en/Home/
2. Basu, R. & Walton, P. (2011). Fit Sigma: A Lean Approach to Building Sustainable Quality Beyond Six Sigma. Wiley. USA. ISBN: 9780470666210
3. Gasik, V. (2013). Designing for Good User Experience with Lean Principles. The Case of Roomforit.com. Pp. 65–75. https://docs.google.com/viewer?urlhttp:// www.cs.tlu.ee/teemaderegister/get file.php?id%3 D210
4. Grove, S., Fisk, R. (1992). The Service Experience As Theater, in Advances in Consumer Research Volume 19 (pp. 455-461), eds. John F. Sherry, Jr. and Brian Sternthal, Association for Consumer Research.
5. Goodwin, K. (2009). Designing for the digital age. How to create human-centered products and services. Wiley Publishing Inc. Indiana. ISBN: 978-0-470-22910-1
6. Hassenzahl, M. (2008). User Experience: Towards an experiential perspective on product quality. http://www.marc-hassenzahl.de/pdfs/hassenzahl-ihm08.pdf
7. Jefferies, E., Tan L., Joyce Y. (2012). WorkPlayExperience: Bringing Drama to Service Design, 1–2. http://design-transitions.com/2012/05/workplayexperience-bringingdrama-to-service-design/
8. Jetter, H-C., Gerken, J. (2010). A Simplified Model of User Experience for Practical Application. Universitet of Konstanz. http://www.inf.unikonstanz.de/ gk/pubsys/publishedFiles/ JeGe06.pdf
9. Kuniavsky, M. (2003). Observing the User Experience: A Practitioner's Guide to User Research, pp. 60–99. USA: Morgan Kaufmann.
10. Liker, J. (2004). The 14 Principles of the Toyota Way: An Executive Summary of the Culture Behind TPS. University of Michigan. Retrieved from: http://icos.groups.si.umich.edu/Liker04.pdf
11. Nielesen, J., Norman D., & Tognazzini, B. (2011). User Experience - Our definition. Retrieved from: http://www.nngroup.com/about/userexperience.html.
12. Ries, E. (2011). The Lean Startup: How Today?s Entrepreneurs Use Continuous Innovation to Create Radically Successful Businesses. New York, New York, USA: Crown Business.

13. Rohner, C. (2008, October 10). When to Use Which User Experience Research Methods. Jackob Nielsens Alertbox. http://www.useit.com/ alertbox/user-researchmethods.html

14. Roto, V., Law, E., Vermeeren A. & Hoonhout, J. (2011). User Experience White Paper. Bringing Clarity to the concept of user experience. University of Helsinki. http://www.allaboutux.org/files/UX-WhitePaper.pdf

15. Roto, V., Lee, M., Pihkala, K., Castro, B., Vermeeren, A., Law, E., Väänänen-Vainio-Mattiala, K., Hoonhout, J. & Obrist, M. (2012). All about UX. Retrieved from: http://www.allaboutux.org/

16. Seddon, J. & O'Donovan, B. Retrinking Lean Service.://www.processexcellencenetwork.com/ downloadSecure Content.cfm?ID=10

17. Womack, J.P & Jones, D. (1996). Lean Thinking. http://www.google.fi/books?id=2eWHaAyiNrgC&hl

Co-Designing for Life: three examples of human-driven design practices for sustainable services

Veikko Ikonen*, Marketta Niemelä*, Minna Kulju*, Antti Tammela*,
Tuomo Kivinen*, Maarit Wallenius*, Heidi Korhonen*, Eija Kaasinen*
VTT, Tekniikankatu 1
33101 Tampere
*{firstname.lastname}@vtt.fi

Maria Alaoja
University of Tampere, Kalevantie 4
33014 Tampereen yliopisto
maria.alaoja@uta.fi

ABSTRACT
In this paper, we describe experiences of various projects when exploiting a collaborative design approach for the development of new services or products. Early user involvement in the product development process is nowadays generally accepted and standardised [7]. Also broad stakeholder involvement, ecosystem and living lab approaches are introduced for more contextual, need-driven and holistic development of service system. [6] However, the methods and techniques for accomplishing this may vary widely. Various methods naturally fit for certain special purposes but also customizing of methods for special needs and particular contexts is very important. Broad stakeholder involvement and commitment for lengthy development process needs to be taken into account when planning design process. In addition co-design and authentic, active user participation to design, calls for the new approaches, methods and tools and even renewal of the traditional model of user centered design.

Keywords
Human-Driven Design, Ubiquitous Computing, Co-Design, Innovation Showrooms, open innovation, stakeholder-based design.

INTRODUCTION
Today, human-centred design is quite an established practice for designing products and services so that forthcoming users are represented in the design process [7]. Human-centred design starts once the decision to design a certain kind of service has been made. To increase the users' role in design and innovation, we should increasingly involve them in deciding what is needed and what kinds of services should be designed for them and with them. Kanstrup and Christiansen [12] describe this change as changing the user's role in design from a victim who needs support to a valuable source of inspiration. But it is still uncertain whether a product or service development team will actually commit themselves to employing potential users in designing the forthcoming product or service as early as possible.

Reasons to this arise from the increasing cost of the early product development phase and the difficulties in finding and understanding correctly the available methods. The most often used methods for early design phases have been interviewing or focus groups. In addition methods for design sessions or for user requirement capturing have been adapted from other disciplines (e.g. ethnography, applied anthropology and participatory design) to better understand the actual usage situation of potential users with a new technological application [e.g. 1, 4, 8, 13].

The foundations of HCI-related participatory design can be traced to 1970s Scandinavia, from where it spread to other parts of the western world. User-centred design (UCD) and participatory design have shared many ideas, techniques and methods, but, in a way, UCD has in the past been more technology-driven, focussing on laboratory testing and finding so-called interface-related usability problems. Participatory design, on the other hand, has emphasised some ideological arguments (e.g. democracy) and given a more holistic view to system development [2,5]. For the past twenty years, and partly because of the launching of Ambient Intelligence and living lab approach, the HCI community and UCD methodology have shifted towards a more holistic view of involving users in the product development process. Besides so called traditional methods of design sessions and user involvement, innovative participation methods have been introduced.

One reason for updating methods and approach has been previously mentioned changing role of users in service design [11,12]. Instead of passive research subjects, they are seen as active co-designers and content creators. Users are acknowledged as best experts in their everyday lives and therefore have great potential as sources of innovation. User participation can affect the success of services directly by better quality, fit to needs and faster innovation speed.

The effects can also be indirect such as more customer-centred image, customer-driven organizational culture and increased motivation of employees or other stakeholders.[9,11]

In traditional human-centred design, only small numbers of users have been involved in the design activities. New methods are needed to reach the masses of potential innovators. Computer-supported methods for co-creation with users are one solution [10, 1l]. New kinds of face-to-face collaboration methods are also needed. In the following we describe our experiences from methodological perspective in various projects.

CO-DESIGN WITH EXPERTS: DESIGN PROCESS FOR P4WELL

A novel service concept for pervasive and personal psychophysiological wellbeing and recovery management for working age-citizens based on stress, sleep, and exercise (P4Well) was aimed to be designed and tested in this reported project. The concept was intended for personal and independent use, and it combined psychological mini-intervention, physiological knowledge, and up-to-date wellness technologies. The project was launched at the beginning of 2008 and finished by the end of year 2009.

The project objectives were defined as:

1. To develop service concepts suitable for interventions for better management of daily load, recovery and stress in working age citizen. Intervention concepts should include concepts primarily or even solely initiated and used by the citizen him/herself, and those offered or supported by service providers such as occupational health care providers or health clubs.
2. To develop mini-intervention designs to empower the citizen towards better recovery from daily stress as well as coping methods to reduce stress reactions. The methods should support individual and personalized strategies for coping.
3. To further develop technology tools for integrated means to assess, measure and monitor level of recovery from the physical and psychological load or stress caused by daily life. Methods should be simple, usable, acceptable and affordable. They are primarily based on personal diaries, monitoring tools, and web-based integration and teleconsultation tools.
4. To develop tools for service orchestration supporting remote (electronic) delivery of both assessment, intervention, and follow up of the citizen.
5. To validate and demonstrate the viability of the concept with real users in intervention studies.

The essential research challenges were:

1. Development of new innovative service concepts which utilize efficiently novel technologies in a multidisciplinary team with experts from many different organizations and backgrounds.
2. Integration of technologies and information from different methods to allow its utilization in interventions as a feedback to an ordinary citizen – either supported by a coach or managing independently his/her health.
3. Design and implementation of field studies to demonstrate the validity of the concept.

Design Process for service concepts

In the design process of the P4Well concept, cross disciplinary working methods were used to design the concept requirements and to discuss usage scenarios of the service. The design process was carried out in a series of expert group (EG) meetings. The EG consisted of experts with various backgrounds (e.g., in engineering, physiology, psychology, and business) from participating organizations. The goal of the EG meetings was to design and develop requirements and a general description of the functionalities and elements of the concept. The approach for the concept creation was need-driven. The primary assumption was that the necessary personal health system (PHS) technologies are mainly available, but the immature applications, services and business models are the main hindrances for the wide spread utilization of PHS.

Efficient yet inspiring working group atmosphere with the experts from the different fields was targeted to be created in the EG meetings. In order to achieve this, collaborative working group methods were emphasized to collect opinions from all the stakeholders to create genuine cross-disciplinary discussions. Most of the organized meetings lasted one working day. During the design process, the EG met six times, with varying themes from the general requirements of the concept to scenario building, ethical and juridical issues, technical functionalities, and related business models.

List of Seminars

[Seminar 1] 20.2.2008, Helsinki, first expert group (EG) meeting: brainstorming for new ideas, examining potential target groups, ethical issues, and identifying basic elements for the service concept. A dialogue workshop method was used: punctilious manuscripted group discussions under predefined themes and groups. The goal was to get different perspectives from the various stakeholders for the concept development in a deliberative manner. 15 EG members participated in the meeting.

[Seminar 2] 11.3.2008, Helsinki, second EG meeting: user stories, use cases, and their analyses, as well as pathways of usage. A preliminary task and scenario-pair writing methods were performed. We organized one hour writing

sessions as a pair work for P4Well user scenario creation. 13 EG members participated in the meeting.

[Seminar 3] 31.3.2008, Helsinki, third EG meeting: legal and ethical issues. In this meeting, we applied a conventional team work method. The aim of the meeting was to find out the most critical ethical as well as juridical issues related to the service concept. 16 EG members participated in the meeting.

[Seminar 4] 22.4.2008, Helsinki, fourth EG meeting: an intermediate summary report of the concept work was launched, and we revised the results of our previous meetings. The goal was to investigate the material gathered for the concept and (re)direct the remaining concept development work according to the made conclusions. 16 EG members participated in the meeting.

[Seminar 5] 13.5.2008, Tampere, fifth EG meeting: technology concept, functions, and critical requirements. In this meeting, a so-called card-sorting game was "played" in pairs. The aim of the meeting was to prioritize and categorize most relevant features for the concept in different usage phases using the collaborative, as well as entertaining, approach. 16 EG members participated in the meeting.

[Seminar 6] 2.-3.6.2008, Summassaari, sixth EG meeting: business models, value chains, and concept review. This meeting consisted of pair and group works with the stakeholder cards from the previous meeting with the potential business platform. The goal was to identify potential business models for the P4Well service concept and also study potential opportunities and threats related to the analyzed models. 19 EG members participated in the meeting.

Expert group meetings succeeded well, with high commitment and active participation from partners. The beginning of the project is critical phase: you have to convince partners that the participation to the design sessions is useful. We took very seriously planning of the workshops and conducting of them: we exploited quite much resource to the actual planning but even more to the analysis and follow-up of the workshops. We took very seriously also participation to the workshops: we e.g. prohibited the use of computer or mobile phone during sessions. This was something extraordinary for participants who were used to check and send emails during meetings. However participants accepted this rule and due this everyone could really concentrate on meeting and enjoy fully of participation.

Outcome of design process was commented to be very useful for further development of the concept. The actual user participation to concept development was done in next phase. Users join in the design process in the form of scenario evaluations, design sessions and field trials. In the end, project itself was very successful and partners of the project were happy with the work done in the project. The concept is still under development while partly it has been already implemented to the some service portfolios of the partners. However, was there something missing from early concept definition phase? Maybe stronger interaction with actual users in first phase could have helped to make concept even better and boosted the development of concept to implemented service faster?

CO-DESIGNING WITH USERS- IHME

IIHME research project takes place during 1.1.2009 - 31.10.2010 and it was segmented into three minor subprojects that offer distinct perspectives to examine methods and applications of ubiquitous computing from human-driven perspective: Developing and deploying new deliberative methods for designing services and environments (DELLU); concepting, developing and introducing environments for open co-design of new services and products (UBIT); and developing services and products by paying attention to the preferences and needs of consumers with novel kinds of profiling solutions (TASSUA).

The main objective of UBIT subproject was to evaluate and develop further the concept of the new innovative experimental environments that work as a platform for testing and demonstrating new ubiquitous computing ideas, as well as, to enable possibilities for users to participate in both designing and executing new environments and services. Creating interactive and experimental concepts of co-design in an entertaining way aims at generating additional information and understanding on producing experimental services and applications that benefit the everyday life of people. In next chapter we'll describe in more details the IHME –environment, main result of UBIT –subproject.

IHME –space

Launching Ihme environment was globally the first trial to test and develop further the idea of an open public co-design environment. Ihme was created to enable an open, low threshold environment where users can freely visit and with guidance experience and test new ubiquitous technology application and share their opinions.

The idea of the IHME -space was to develop experimental environment based on the strong interaction and participation of the users. The users were given the opportunity to affect, and to take part in the innovation process as co-designers for new technology and services.

VTT Ihme environment was located in the Ideapark shopping centre (Lempäälä, Finland) in a 61-square-metre facility. In the facility users were able to try out applications based on ubiquitous technology. Applications in the show room included virtual games on the pressure-sensing floor (Pomppaa!), as well as games utilizing augmented reality (Dibitassut and Mobiililogot). Users

were also able to visit travel destinations virtually (VirtualTravel) and get acquainted with 3D autostereoscopic display and SmartBoard interactive whiteboard.¨

Figure 1.Front view to the IHME -space.

The staff included two employees dedicated to guide on learning the main principles and usage of the applications in the show room. After a guided tour the users were asked to fill out a short questionnaire concerning their preferences and ideas of further developing of the applications; and, as well, on how they experienced both visiting the show room and participating on designing future technologies in the offered shared environment.

After being accustomed and having experimented the applications visitors were asked to fill out a short questionnaire concerning their opinions and experiences on participating in co-designing process. Besides the written feedback collected on PCs, the users also gave direct oral feedback and developmental ideas to the staff. Actual notes were not written by the employees but the users were encouraged to write down their propositions on the questionnaire form.

Questionnaires used in the survey included two multiple choice questions which utilized a Likert scale for measuring users opinions and experiences about the user-friendliness and simplicity of new technological applications and services. These questions were created to map out a technological background and basic knowledge of the typical user visiting the IHME environment. Users were also asked to point out the most appealing applications.

Questionnaires section for open questions was designed to create an overview of user-experiences of visiting an open co-design environment and participating on ideating new services and products. The users were also asked to ideate new ways of applying technology presented in the show room; and to describe the features of the applications that made them either interesting, or not appealing. The users were, as well, asked to describe their stance on participating on developmental ideating of future technologies.

RESULTS

IHME space reached a great number of visitors (c.a. 2500). 355 users filled out the questionnaire. Majority of the visitors in the show room were children aged approximately from 7 to 12 years who, however, were only partially represented in participating in the questionnaire. According to questionnaires answered the average age of users visiting the show room was 35 years and the distinctive majority, 65%, of the visitors were men.

The VirtualTravel developed in IHME -project (presented and developed during the summer in Ihme environment) was the most appealing application (figure 1). Interactivity, entertaining factor, novelty, innovative visual representation, presence of a sound feedback, possibilities for further development and for a broad applicability were mentioned as the reasons making certain applications more appealing than the others (VirtualTravel, Smartboard). As a common reason, why other applications didn't make an impression, users stated that they couldn't see the practical benefits or the applicability was not easily noticeable (Dibitassut, Mobiililogot). Other commonly listed reasons were lack of novelty (Pomppaa!) and a game-like appearance of the applications (Pomppaa!, Dibitassut, Mobiililogot).

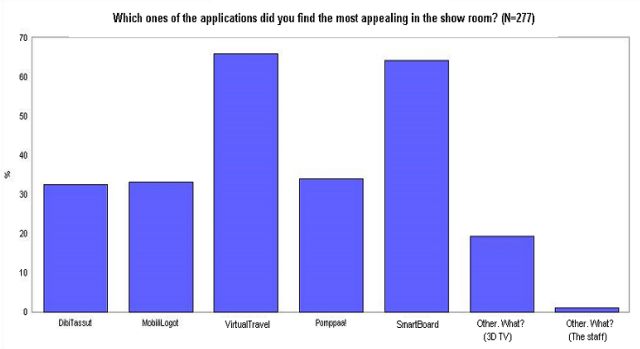


Figure 2. Distribution of the applications by their attractiveness.

Tech-savviness of the users visiting Ihme environment was measures on a five-level scale mapping out if users found new technology rather complicated and difficult (5) or easy and simple (1) to use. The distribution shows that 12 users (n=277) found new technology difficult and complicated (value 5), 63 users stated it to be rather more difficult than simple (value 4), 64 users didn't find it either difficult or simple (value 3), 81 users said it to be rather more simple than difficult (value 2) and 59 users experienced new technology to be simple and easy to use (value 1).

New innovative ideas of applying technology in other contexts, as well as the ideas for the further development of the applications were expressed in 120 feedbacks. Majority of the feedback given in this section included innovative and creative ideas of applying technology in new context improving peoples everyday life.

VirtualTravel application gathered the most feedback. Users invented numerous ways to utilize the application. Utilization as a tool for visiting museums, concerts or and for virtual shopping were mentioned frequently. Users figured possible benefits also for informational and

commercial purposes in public environments and for usage as stimulating equipment for elders, invalids and sick people in hospitals or old people's homes. Also the idea of using the application as an implement for lecturing and other educational purposes (for example teaching biology or geography) was pointed out. For further development of the VirtualTravel application users brought up ideas to have live material, a touch screen and 3D –features (both visual and audio) included to the application, a possibility to use the application online was also commonly expressed.

Visitors were also asked to state their stance on how they would feel about participating in designing new technologies and services. Out of the 175 answers given to the question 120 (69%) showed a positive response. Participating was seen as a useful and important, not to mention fun and interesting, way for stating an average users point of view, and also, necessary to developing user-friendly products and services. Convenient ways of participating listed by the users were short group interviews and conversations, online questionnaires, entering a publicly open show room and participating as a test user of new products and services. The main reasons why users were negative about taking part in the development process were not having enough time or interest, or users saw themselves not tech-savvy enough.

Visiting IHME space was reported as a positive experience by all the users answering the questionnaire. Positive user feedback was given concerning the opportunity to participate and experience new trends of ubiquitous technology which visitors found useful, public appearance of the VTT, easily approachable location and low threshold to enter the environment and to take part, expert knowledge and friendliness of the staff, and the children-friendliness of the Ihme environment. Negative user feedback concerned complains about presented technology being outdated, applications experienced as either too simple or not easily applicable in any useful way.

This human-driven design method supplied a contrast to the predominant trend of device-driven design IHME environment presented a globally unique trial for new kind of co-design approach: showroom and living lab in same context. It proved ability to enable a low threshold user research environment, obtaining feedback on innovative technology from ordinary people and other stakeholders in an open public environment. VTT IHME environment' reached a great number of visitors and enabled a large amount of direct involvement and feedback from users in designing products and services. Direct contacts and interest to co-operation with potential partners/companies were obtained and great visibility in media for innovative research results was provided .Not only did the users find it informative and important to participate on the evaluation - every single one of the visitors found it entertaining and fun to participate. The trial provides indispensable knowledge for the future projects of designing co-design environments.

CO-DESIGNING IN CONTEXT: SHAPING MARKETS FOR SUSTAINABILITY – SHAPE

In Shape –project we take further steps in developing and implementing of co-creation methods and approach. Our cases with companies are focusing on travelling, furniture industry, local food in supermarket, sustainable amusement park in big shopping centre. We have utilized expert meetings, user interviews and questionnaires to get basic knowledge of these particular cases in context. The next steps of the study are focusing on real co-creation either in virtual or physical co-creation platforms or spaces. One example of our own development in tools for co-design is Visual -IHME. Another example is our co-design activities arranged in conjunction of travel and housing fairs.

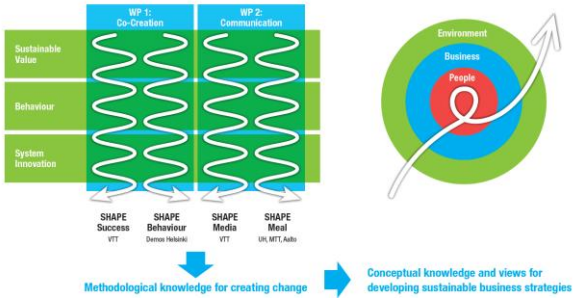


Figure 3. General approach of SHAPE

Co-designing in Fair environment and in Train

Our aim was to involve various stakeholders to the designing of sustainable services in context where participants are already tuned in to the right mood: looking for travelling experience.

Domestic Travel Fair was held on 13-15 April in Tampere Exhibition and Sports Centre at the same time with four other fairs We participated in Domestic travel fair with Tampere Region Economic Development Agency Tredea at the Tampere stand

Main purpose was to study: 1) fair visitors' opinions on sustainable travelling and factors important to them while travelling; 2) Fair as an environment for ad hoc short interviews.

The study sessions consisted of an open question: what sustainable travelling means to you and what comes to your mind on the words "sustainable travelling" and a short questionnaire (1xA4) for the participants to fill. The study sessions took appr. 5-10 minutes. Participants were selected randomly among visitors, mainly from the main entrance hall (110 participants). The questionnaire was also available on the internet (61 participants, but the questionnaire on the internet didn't include the open question on sustainable travelling due to technical reasons.

It is fair to say that the context (travelling fair) made it easier to people to think and give their opinions. Also to the designers, researchers and other involved doing research in context was enjoyable experience: it was easy to ask people

to participate for a shorter or a longer discussion depending on time of the people could spend. We focused this spring on "slow travel" – experience which was identified as a one of the potential features for sustainable travelling. One of our experiments in this focus area was to conduct innovation sessions and user interviews on train while travelling from Helsinki to Rovaniemi (so called Design Train). The experience was encouraging: expert sessions in their own cabin were successful, feedback was very positive from this group. Also user interviews were mainly commented as very inspiring experience: people were willing to spend a few moments during their travelling for interview and discussion was also easily related to the design issues – slow travelling.

Visual-IHME

The Visual IHME end-user interface shows a "window" to a spherical panorama picture that covers an environment in all directions from the point the picture is taken. The window can be turned around with a mouse, or with finger if a touch screen is in use. The user can move from one panorama picture to another by clicking an arrow button (Fig. 1 down left). Sound can be added to the panorama to increase realism. Also other types of files such as images, photographs, and texts can be attached to the panorama picture.

Figure 4. Distribution of the applications by their attractiveness.

The current set of co-creation tools include comment/discussion boards (either pinned to a location or positioned aside the image) with a "thumb up/down" voting function, questionnaires and polls. Within the current system, any user is able to start a new discussion topic but only users with editor rights can create questionnaires. The gathered data is stored in a database in the Visual IHME software. The data can be exported and analysed in Microsoft Excel format. In the future, more co-creation tools with an elaborated user right system and tools for data analysis will be developed.

We carried out a small scale end-user evaluation of Visual IHME with 10 adult volunteers of different ages (seven female, three male) on February 2013. The participants were individually given a brief demonstration of the platform and its co-design features, then filled a questionnaire and were shortly interviewed.

Visual IHME got positive feedback from all of the participants. Nine participants said that would use this kind of tool to share and develop ideas with others. The most valued features were "thumb up/down" (eight participants of the ten thought that this function is necessary in the interface), discussions (7/10), ability to create new discussion topics (6/10) and questionnaires (6/10).

In general, the participants thought that the role of the platform is to support collecting ideas from consumers and elaborating them together with designers. We received some ideas how and where Visual IHME could be applied to, for instance: when designing usage of land, traffic or telecommunications network, and when designing for equality and accessibility (e.g., accessibility of a city with the disabled inhabitants as one stakeholder group in the co-design process).

Each participant found social media services and internet as overall good tools for involving users in design and development, especially when developing new services or technology. However, half of the participants had a conception that with (current) social media services they are *not* able to influence issues that are important to them. In line with Dourish (2010), we take this as a hint that even in the modern world penetrated with communication technologies, there is still plenty of space to develop virtual or technology mediated platforms that really can empower people to influence the conditions they live in.

DISCUSSION

Design of future ICT and sustainable services and products that are mainly based on ICT, calls for new design methodologies for the greater acceptance among presumable user groups and justification among all relevant stakeholders. Furthermore the factors affecting to the design decisions of future services and applications are numerous, various and in many cases frequently in conflict. Today user experience studies are conducted in order to get better effectiveness, efficiency, and user satisfaction and user acceptance for new products and services. Even though the design approach is called human-centred or user-centred the design of new products and services has been quite technology or market driven in Information and Communication Technology (ICT) business. Instead of putting technology or market to the core of design process and product development the human needs and values should form the fundamental basis of design.

Human Driven Design (HDD) refers to the design approach which broadens the perspective from focused product or service development process model to the more holistic design perspective. Co-design as a lower level methodology furthermore broadens the scope and role of involved participant groups in the actual design process. Authentic co-design aims to give voice and influence to all

relevant stakeholders and in that sense it links closely to the deliberative design approach. HDD includes assessment of critical issues (i.e. social, ethical) of design process and artifacts while tradeoffs in design have to be well justified. The ultimate goal of HDD approach is to accomplish design framework which will empower all stakeholders when designing their everyday service environment or context of work.

ACKNOWLEDGMENTS

We would like to thank P4WELL, IHME and SHAPE – project partners for their contribution. Without their involvement projects would not have been possible to realize. We also like to express gratitude to Tekes – National Funding Agency for Innovation. Last but not least we give enormous compliments to the people who have voluntarily given their valuable contribution to the projects. Hopefully you have had as fun as we had doing these projects!

REFERENCES

[1] Beyer H, Holzblatt K (1998) Contextual design: defining customer-centered systems. Morgan Kaufmann, San Francisco

[2] Clement A, van den Besselaar P (1993) A retrospective look at PD projects. Commun ACM 36(6):29–37

[3] Dourish, P., 2010. HCI and environmental sustainability: the politics of design and the design of politics. Proc. of the 8th

ACM Conference on Designing Interactive Systems (DIS 2010). Aarhus, Denmark, 1-10.

[4] Ehn P (1989) Work-oriented design of computer artifacts. Erlbaum, New Jersey

[5] Grudin J, Pruitt J (2002) Personas, participatory design and product development: an infrastructure for engagement. In: Proceedings of the 7th biennial particpatory design conference (PDC 2002), Malmo¨, Sweden, June 2002, pp 144–161

[6] European Commission (2009). Study on the potential of the Living Labs approach Including its relation to experimental facilities for future Internet related technologies, final report, study performed by Altec, Version 1.1, 9th March 2009

[7] ISO 13407 (1999) Human-centred design processes for interactive systems. The International Organization for Standardization

[8] Hughes J, King V, Rodden T, Anderson H (1995) The role of ethnography in interactive systems design. ACM Interact 2(2):56–65

[9] Ikonen, V. 2009. Towards Empowering practices. Ambient Intelligence and Smart Environments, Vol. 1, Ambient Intelligence Perspectives, Selected Papers from the First International Ambient Intelligence Forum 2008. Mikulecky, Peter et.al. (ed.). IOS Press. Amsterdam, 129 -136

[10] Jensen, J. L. (2003). Virtual democratic dialoque? Bringing together citizens and politicians.. Information Polity, 29-47.

[11] Kaasinen, Eija; Koskela-Huotari, Kaisa; Ikonen, Veikko; Niemelä, Marketta; Näkki, Pirjo. Three approaches to co-creating services with users.1st International Conference on Human Side of Service Engineering, San Francisco, USA, 21 - 25 July 2012. Conference Proceedings 4thInternational Conference on Applied Human Factors and Ergonomics (AHFE) . USA Publishing. Stoughton, Wisconsin, USA (2012), DVD publication

[12] Kanstrup, A. M . and C hristiansen, E .: Selecting and evoking i nnovators: Combining democracy and creativity. In: Proceedings of the 4th Nordic conference on Human-computer interaction: changing roles. (2006).

[13] van Willigen J (1993) Applied anthropology: an introduction. Bergin & Garvey, Westport

Impact of Sensory and Collaborative Contexts on Usability Evaluation Methods Resources

Ebba Thora Hvannberg
University of Iceland
Dunhaga 5
107 Reykjavik, Iceland
ebba@hi.is

ABSTRACT
This paper describes the evaluation of two applications which are different in their sensory contexts and collaborative contexts. For each application, five method resources are analysed which were used for their evaluation to see what impact the contexts had on the method resources. The analysis concludes by discussing if the transfer across the contexts was successful. Formalising the transfer between contexts, the paper proposes three types of transfer function genres, generalisation/specialisation, adoption accompanied with adaptation and transfer between domains.

INTRODUCTION
With the ever increasing variety of contexts in which usability evaluation is carried out, researchers and practitioners propose modified or new methods that are tailored to specific contexts such as virtual reality to name one. To aid practitioners in selecting from the vast flora of methods, researchers have proposed frameworks with characteristics that become the basis for the method selection. Examples of such characteristics is the objective of the evaluation, e.g. precision (why), the outcomes and metrics of the evaluation (what), whether the evaluation is summative or formative (when), and which types of methods e.g. questionnaire or observation, to choose (how) [1, 2]. In addition to the above characteristics as a basis for selection, it has been well recognised that the resources available, e.g. people, time and money, may determine the type of evaluation applied [1].

Studying and designing new evaluation methods, researchers have proposed evaluation frameworks that can be used to select a certain method. To further narrow the scope of such frameworks, some have been targeted to specific types such as collaboration. The motivation or the process of transfer of methods within those frameworks is seldom seen. Methods evolve stimulated by some need and experience. Although, transfer has not been investigated extensively, there is clearly interest in learning more about it. For example, Antunes et al. [1] looked at 48 papers on reusable evaluation strategies of collaborative systems and saw some identifiable patterns of transfer between methods. The first is an adaptation of single-user evaluation methods to the context of collaborative systems. The second is an assimilation of methods and tools from other fields beyond technology development, e.g. ethnography. The third pattern identified by Antunes et al. [1] was that while the early methods were targeted for specific measures carried out in controlled situations, latter ones were tailored for more complex and broader contextual concerns.

Woolrych et al. [3] concluded that instead of seeing a method as a fixed entity, it would be more useful to view an evaluation method as having a number of resources being any pre-existing re-usable component of design or evaluation that can contribute to the formation of a method through practice.

This paper aims to analyse how usability evaluations are transferred between contexts with respect to five method resources. The transfer of method resources will be matched against three forms of transfers. The hope is that the study will increase our knowledge on transfer function.

'OLD' METHOD-RESOURCES IN NEW CONTEXTS

Two training applications
In this case study we will describe how usability evaluation has been carried out on two applications which are both training software. One is meant to train students in mathematics and the other to train crisis response and management. The former is targeted to blind students and the latter targets workers in several crisis management sectors, from fire fighters, police, rescue teams and medical personnel, responders and commanders, professionals and volunteers.

The mathematics application, Mathematics Cane, is a single user application, but the Crisis Management Training software (CMT) is a collaborative application since the work requires teams to coordinate and communicate on tasks. In both cases, the user population is specific and narrow, defined by their sensual abilities in the former case and work domain in the latter case. What they have in common, besides being training software, is that they apply non-visual user interface technologies which are meant to aid the user in completing the tasks. In

the case of the training software in mathematics for blind students, a haptic peripheral was used augmented with sounds, but in CMT soundscapes and voice communication characterised the virtual environment.

Method resources

For the purpose of this comparison, we have chosen to focus on five types of method resources [3]. The first two, participant recruitment and task selection, are generic for user testing. Training objective selection, the third method resource, is specific to training software and is a special case of evaluation criteria. The fourth one is data capture which we think is relevant when investigating innovative user interface technologies. The final method resource on our list is control of experiment:

- Participant recruitment
- Task selection
- Training objective selection (evaluation criteria)
- Data capture
- Control of experiment

Before describing the two application contexts, we will say a few words about the above method resources. Participant recruitment is about how users are invited and selected to take part in the usability evaluations. Invitation can be via e-mail, written letter, through a social network or from another population of users. The recruitment can be obligatory or optional and the selection can be according to criteria which are of varying restrictions.

Appropriate task selection and task description has been long considered vital [4, 5]. What is thought to be most important is that it matches the user group. Another criteria for selection is the scope and the span of the tasks, i.e., whether only a small fraction of the tasks offered by the application is chosen or a large one. The task selection will depend on the overall objective of the experiment. Sometimes, evaluations are carried out to test specific parts of a system which were found to be unsatisfactory in previous evaluations. The task selection will determine the length of the experiment and hence depend on how much time participants can devote to the experiment.

Whereas task selection results in a medium-grained set of tasks that users are instructed to carry out, applications can have an overall objective, as in the case of training software to train students in a skill or a competency. In contrast to evaluating if a user can complete a task it is desirable to see if he/she has achieved the objective of having this skill. This overall objective may or may not be considered a part of usability evaluation, but since evaluations of usability usually include effectiveness, it is natural to consider it.

Data capture means any tool or method with which we seize data from the usability evaluations to extract qualitative or quantitative data. What data is captured depends on the instruments applied, e.g., for heuristics evaluation, the main objective is to uncover problems. Data is captured manually by the expert but sometimes screen is captured of the execution of the task automatically so that the expert or a second expert may review the data and verify or carry out an independent review. Several other data capturing methods exists such logging of events, eye tracking and capturing physiological metrics.

Above, we discussed task selection and granularity of tasks. A related method resource is how these tasks are put in front of the users and what freedom they have to carry them out. In a tightly controlled experiment the tasks may be fine grained and put in front of the users in a systematic manner, but in a loosely controlled experiment, the users will be given overall problems which they are to solve. Thus, we note that the individual resources are not discrete but coupled together make up an integrated approach for an evaluation to take place.

Contexts

There are different contexts in which design and evaluations are carried out. These contexts are determined by diverse stakeholders, development methods, technologies, maturity of companies, access to users, bespoke or off-the shelve software, to name a few.

We have chosen to describe two contexts which show the variability between the two applications under consideration (see Table 1). The contexts are the sensory context and the solo/collaborative context. The sensory context of the Mathematics application was haptic, sound and voice synthesis to accommodate blind students, and the sensory context of the CMT application was graphics, sound to accommodate navigation, situation awareness and reality of the noisy environment, and voice. The solo/collaborative context of the Mathematics cane was solo and collaborative for the CMT.

Description of Original Application Context: Single User Training Software for Mathematics with Haptics, Sound and Voice Synthesis

The Mathematics Cane is a software application developed by a master student,

Table 1 Two applications characterised by the sensory context and the solo/collaborative context

Appli-cation	Sensory Context	Solo/Collaborative Context
Mathe-matics	Haptic, Sound, Voice Synthesis	Solo
Crisis Manage-ment	Graphic, Sound, Voice	Collaborative

Magni Þór Birgisson [6], aimed to help young blind students to learn about 2D polynomials. It accepts polynomial equations and allows students, applying a haptic peripheral, to feel the shape of the polynomial. In addition to the kinaesthetic aid, a student can choose to have sound added when navigating the polynomials with the haptic peripheral. Furthermore, a grid can be superimposed with an extra force (like a tic) felt when passing over the grid crossings. The grid is to make it easier for the blind students to count as they examined the polynomials with the haptic peripheral. A research study was carried out to see which modes, haptic, sound or voice synthesis helped the students. In the remainder of this section, we will address each of the five method resources, starting with participant recruitment.

Participants were recruited from a group of blind and seeing students, six in each group, age 15-25. Since the user population of blind students in this age range is small, seeing students were included, but blindfolded. Furthermore, the aim was to see if there was any qualitative difference between the seeing and the blind students. There was no minimum threshold on mathematics competency, but in retrospect this may have been advisable.

For the tasks, four polynomials were selected with a variability of difficulty, from easy polynomials to challenging ones. Each polynomial was presented to users with and without sounds and grids with the objective to explore the different mode's effectiveness.

The training objective of the study was to assess the students' perception on polynomials by asking them to compare the haptically drawn polynomials to embossed ones on paper (3 options given). Hence, only a vague training objective was selected. The evaluation criteria of the research study included research questions on the usability of different sounds, haptic and voice while exploring the polynomials and if the Mathematics cane could be an improvement beyond the current technology.

Data was captured in several ways. An observer noted the time it took users to complete the tasks, intermediate steps of the tasks, such as when the students started to look at the embossed paper for comparison. Right and wrong solutions were recorded, deviations and problems. Using a five point Likert scale, the observer asked participants how easy they thought it was to solve a task, if the sound was useful and the grids. They were asked if they thought that they had solved the tasks efficiently. A software implemented log of users' navigation through the polynomials was stored. In another research study, software was written to visualize the log from the navigation of the haptic movements [7].

Considering the fifth method resource, i.e. the controllability of the experiments, the research study had the tasks put in front of the user one after another, and thus, it was tightly controlled.

Description of New Application Context(s): Collaborative Training Software for CMT with Sound, Voice and Graphics

A training simulator set in a virtual environment has been built for training crisis management. A prototype of the simulator has been implemented to evaluate realistic soundscapes, voice communication and noise from the far end over the voice communication [8]. The simulator is multi-player, configurable with different crisis scenarios that typically occur in the transport sectors, such as in airports or railways. The objective of the training software is to train responders and commanders in triaging casualties, reporting the results in a command hierarchy, making resources available and transporting casualties to hospitals or shelters as appropriate based on the prioritisation of the injuries. Sounds, generated by objects, such as fire trucks, and chatter are implemented in the virtual environment. Three different communication metaphors, radio, mobile phone and face to face is made available in the simulator. The current prototype is limited to two persons speaking at any time over one of the metaphors. The remainder of the section describes the five method resources, participant recruitment, task selection, evaluation criteria, data capture and experiment control.

Six participants were recruited from a local fire- and rescue company to carry out training using the simulator in three pairs. The participants had been attending a course on simulator training software, but not the same as the one that was evaluated. Obviously, they were not recruited from a large population and they may have been self-selected in a way that the more computer-skilled employees may have shown more interest in the course and the study. There were no additional competency criteria, e.g. novice or experienced, or any particular roles, beyond working and receiving training.

Tasks were selected to evaluate the voice communication, perception to sound, navigation in the virtual environment, situation assessment, presence and flow. A scenario was written with a total of 14 tasks for two roles collaborating, an On Scene Commander and a Rescue Coordinator.

The training objective of the simulator is to train 12 competencies, and three of them were specifically trained in the above scenario: to assess alert, communicate and to communicate information. Seven research questions were posed for the research study, mainly on the effect of sounds, noise and the usability of voice communication using different metaphors.

Data was captured by an observer following each of the two users in the experiment and screen and audio captures, thus storing all graphics, sound and voice

communication. Afterwards, all voice communication was transcribed verbatim. Questions on effectiveness, efficiency and satisfaction were asked in a post-test review, asking these questions for each task. This was different than in the Mathematics cane study since doing it for each task would have broken the flow of the scenario.

The 14 tasks in the scenario were put in front of the user one by one by an observer. Hence the experiment was controlled tightly. This was not so successful and could be attributed to the requirement of letting participants engage in the virtual environment instead of being interrupted by the test facilitator saying what tasks to do next.

Stories of Transfer: Triumph and Tragedy

This section describes if the method resources needed to be transformed from one context to another. In an informal study that asked experts about possible transfers of evaluation methods, out of 20 stories we further analysed 12 of them. The outcome of that analysis gave three forms of transfer. First was the **Generalisation/Specialisation transfer**, where there is a generic method (resource) for a broad set of application but it is desirable to specialize it for a certain context. We have many examples of this transfer function, e.g., for Heuristics Evaluation, where the method resource is a list of Heuristics, e.g. specifically for virtual environments [9, 10].

Another example of **transfer is an adoption, often accompanied with some adaptation** [11], where a method resource is adopted from a discipline to be used in a new context. An example was: *"As usability professional designing interactive and tangible children's toys, having found laddering from the marketing domain I need to know what aspects need to be changed to apply the method to evaluate the affective responses to specific attributes of the prototypes."*

A third category of **transfer is between domains**, e.g., from gaming to learning. An example requirement to a method source transfer was "*As a teacher (e-learning)/designer, I want to explore animation techniques from certain games, so that I can increase learner's engagement in exploring content in a reflected way"*

In the analysis below, we will see if the transfer of each of the method resources could be attributed to the above three forms of transfer. Each of the two application contexts (old or new) was characterised by two variables, the sensory context and the solo/collaborative context (Table 1). In our model, each variable can have two values. Although the participants in the two evaluation experiments are quite different in their skills and background, the participant recruitment procedure is similar in both cases and neither the sensory nor the collaborative contexts affected the recruitment procedure itself. In both cases, we set some criteria on whom we would select and asked for local help for the recruitment. A contact at a school helped with the recruitment in the case of the Mathematics and a contact at a local fire- and rescue company for the CMT. Hence, in both cases a facilitator close to the users helped with the recruitment. In both cases it was discovered during the tests that a further pre-screening of the participants for certain competencies would have increased their fit to the desired target group. The collaborative context had a slight impact in that in the CMT case we had to recruit two participants for each experiment. Hence, the transfer of this methods resource was successful and required specialization of the participant criteria which was directed by the skills and abilities of users of the target application domain.

The task selection, in both cases, depended on the desired outcome of the evaluation. For the CMT case the design of a scenario consisting of several tasks proved much more challenging since a realistic scenario had to be written covering the usability of the application and the specifics of the sensory channels (i.e. sounds, voice, and graphics) of the application. This transfer, from simple individual tasks of the Mathematics cane to more complex realistic scenarios including sound perception was partly unsuccessful. In the CMT case some users noted that a task was unrealistic and not according to crisis management procedures. Furthermore, it was hard to select tasks that tested the realism and appropriateness of sounds. Scripted training is bound to produce different results than free flowing open, non-scripted training. In particular, non-scripted testing is not repeatable and since participants are working together in the collaborative scenario they will influence each other's actions [12]. We deem the type of transfer to be an adoption of techniques from other disciplines such as theatre, creativeness and flow [13] [14].

In neither experiment a formal training objective was assessed. The experiment with the Mathematics cane only evaluated if the users could recognise polynomials, but did not evaluate if the software helped the students gain more knowledge of polynomials. Three of the twelve competencies necessary for crisis management were addressed in the latter experiment and were determined by the collaborative contexts. The evaluation criteria for each application were designed for the research study and were heavily influenced by respective contexts. For the CMT and the Mathematics cane the sensory context influenced the evaluation criteria, i.e. sound, noise, voice and haptics but in different ways.

Data capture is the method resource which varies the most across the two contexts. Data was captured with special analytics software in the Mathematics cane but with regular screen capture and audio recording in the CMT. The analytics software made it easier to track navigation

along the polynomials [7], but no such tracking device was available for analysing navigation in the simulator. Instead, observers would need to mark specific places as visited as they reviewed the video afterwards. One of the challenges in evaluating the effect of sound is that it is not possible to track how a user perceives sound. This has to be judged through their behaviour. For example, possibly, a tracking device could help to conclude if sound and graphics had the desired effect, such as observing if a user is not moving close to a burning fire, and moving away from a fire engine before communicating. If sound had the desired effect, it should be noticed in users talking louder. This can be observed by listening to the audio, but it is better if it can be measured objectively with a sound level meter. We conclude that it takes considerable effort to transfer resources for data capture along the sensory context. The collaborative context influences data capture since synchronization is required between partners and it is important to note the time stamps of the users are synchronized. The form of transfer is rather complex, requiring perhaps all three forms of transfer, specialisation, transfer from other domains, such as games, and adoption accompanied with adaptation from the audio domain and eye tracking.

ANALYSIS OF CONTEXTUAL FACTORS

From the analysis of the success and failures of using methods resources across contexts, we conclude that the sensory context had an effect on the task selection and data capture, and the collaborative context had an impact on all method resources except participant recruitment (see Table 2). We only chose to indicate if a context had an impact on a method resource without quantifying the impact. For example, from our analysis of the two studies, we saw that the sensory context would have a large influence on data capture but collaborative context had a small influence on the same resource. We cannot conclude by analysing only one case of transfer, but there may be variability of impact across individual applications in the same context set.

Table 2 Contextual Factors Influencing Method Resources

	Contextual Factors	
Method resources	**Sensory context**	**Solo/Collaborative context**
Participant recruitment	No impact	No impact
Task selection	Yes	Yes
Training objective selection	No impact	Yes
Data capture	Yes	Yes
Control of experiment	No impact	Yes

The transfer of methods resources across contexts can be formalized with the following formula:

Transfer (Old Method Resource, set of Context-Old, Context-New) = Modified Method Resource

In cases where there is no impact as per Table 2 the *Old Method Resource* is equal to the *Modified Method Resource*. This assumes that the different contexts are independent and impact the method resources independently. However, if they are interdependent all the Contexts (*set of Context-Old)* for the source would have to be a part of the input simultaneously. This is to emphasise that when we cross context boundaries from old to new, project teams may use their collective experiences from multiple old contexts. The transfer function does not separate explicit knowledge, e.g. from the literature, as a variable, but it may be justified to have include it as a separate variable rather than tacit knowledge of the context. In addition to changed method resources, it may very well be that the set of method resources change. Thus, one can omit or add a method resource, e.g. participant recruitment, when crossing different values of a context.

It remains to be seen what kind of function the *Transfer* function is. We summarised three such transfer functions, generalisation/specialisation, adoption with adaptation, and transfer between domains. For each method resource we attempted to analyse if the two studies had applied these transfer functions. The fourth transfer function can be transfer across the sensory context as we saw in the example of the data capture where transfer takes place across sensory and collaborative contexts.

CONCLUSION

Technology, needs and development contexts are all factors that evolve rapidly and provide new opportunities to give enhanced experiences and satisfaction to users. This creates new requirements for design and evaluation methods. This paper has shed light on how this innovation of methods takes place. The analysis of methods resources showed that various factors influenced how the method resources were used or reused in each of the applications. Specifically, we looked at the impact of two contexts, the sensory and the collaborative contexts. We conclude that some of the method resources can be almost re-used directly, such as participant recruitment. The participant criteria had to be redefined and the local contact was different, but the main process was similar. The control of experiment was different for the CMT than the Mathematics cane and because of the influence by the solo/collaborative context that method resource had to be redesigned and could not be reused. Other method resources such as task selection and data capture had to be redesigned because of the influence of the sensory context and the solo/collaborative context. Because of the specialised domain of each application, especially the CMT, additionally, knowledge on processes and practices

had to be taken into account when designing the scenarios of tasks for the evaluation.

Furthermore, the study showed that the method resources can be interdependent. For example, a detailed task scenario will require a high level of experimental control that produces different results than a non-scripted scenario.

In studying how methods transfer from one context to another we build new knowledge upon existing practices. While transfer of knowledge [15] is a powerful tool which has been used in many settings, it is challenging since two contexts need to be studied and the motivation for the transformation between them is hard to discover, what is the input, output and the transfer function which is most complex. This paper has given an example of such a study, where we have analysed evaluations of two training applications and looked at their different sensory and collaborative contexts and their impact on the methods resources. This has been only one of many examples needed to understand the transfer of methods resources across contexts.

ACKNOWLEDGEMENT

The author would like to thank the anonymous reviewers for valuable and constructive comments.

The research leading to these results has received funding from COST Action IC0904 and the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement no. [FP7-242474].

REFERENCES

1. Antunes, P., Herskovic, V., Ochoa, S.F., Pino, J.A.: Structuring dimensions for collaborative systems evaluation. ACM Comput. Surv. **44** (2012) 1-28
2. Law, E.L.-C., Hvannberg, E., Vermeeren, A.P.O.S., Cockton, G., Jokela, T.: Made for Sharing: HCI Stories of Transfer, Triumph and Tragedy. CHI 2013 Extended Abstracts. ACM, Paris, France (2013)
3. Woolrych, A., Hornbæk, K., Frøkjær, E., Cockton, G.: Ingredients and Meals Rather Than Recipes: A Proposal for Research That Does Not Treat Usability Evaluation Methods as Indivisible Wholes. International Journal of Human-Computer Interaction **27** (2011) 940-970
4. Alshamari, M., Mayhew, P.: Task design: Its impact on usability testing. Internet and Web Applications and Services, 2008. ICIW'08. Third International Conference on. IEEE (2008) 583-589
5. Sears, A., Hess, D.J.: Cognitive walkthroughs: Understanding the effect of task-description detail on evaluator performance. International Journal of Human-Computer Interaction **11** (1999) 185-200
6. Birgisson, M.T.: Stærðfræðistafur. Computer Science, Vol. MS. University of Iceland, Reykjavik (2008) 133
7. Jónsson, F., Hvannberg, E.T.: Using a Log Analyser to Assist Research into Haptic Technology. In: Holzinger, A., Miesenberger, K. (eds.): HCI and Usability for e-Inclusion, Vol. 5889. Springer Berlin Heidelberg (2009) 523-531
8. Rudinsky, J., Hvannberg, E.T., Helgason, A.A., Petursson, P.B.: Designing soundscapes of virtual environments for crisis management training. Proceedings of the Designing Interactive Systems Conference. ACM, Newcastle Upon Tyne, United Kingdom (2012) 689-692
9. Hvannberg, E.T., Halldorsdottir, G., Rudinsky, J.: Exploitation of heuristics for virtual environments. Proceedings of the 7th Nordic Conference on Human-Computer Interaction: Making Sense Through Design. ACM, Copenhagen, Denmark (2012) 308-317
10. Sutcliffe, A., Gault, B.: Heuristic evaluation of virtual reality applications. Interacting with Computers **16** (2004) 831-849
11. Shepherd, I.H., Bleasdale-Shepherd, I.: The design-by-adaptation approach to universal access: learning from videogame technology. Univ Access Inf Soc **10** (2011) 319-336
12. Hey, E.: Case Study: Conducting large-scale multi-user user tests on the United Kingdom Air Defence Command and Control system. Journal of Usability Studies **1** (2006) 121-135
13. Jin, S.-A.A.: "Toward Integrative Models of Flow": Effects of Performance, Skill, Challenge, Playfulness, and Presence on Flow in Video Games. Journal of Broadcasting & Electronic Media **56** (2012) 169-186
14. Nakamura, J., Csikszentmihalyi, M.: The concept of flow. Handbook of positive psychology (2002) 89-105
15. Argote, L., Ingram, P.: Knowledge Transfer: A Basis for Competitive Advantage in Firms. Organizational Behavior and Human Decision Processes **82** (2000) 150-169

A Cross-Team Collaborative Evaluation of a CRM System

Marcin Sikorski
Gdansk University of Technology,
Faculty of Management and Economics
ul. Narutowicza 11/12
80-233 Gdansk, Poland
Marcin.Sikorski@zie.pg.gda.pl

ABSTRACT
This paper presents an analysis of an usability evaluation a CRM (Customer Relationship Management) performed by a team composed of external usability experts jointly with a CRM staff. The evaluation process differed from a classical scheme known from former projects, including some new elements resulting from a specific context of this study. These novel elements resulted in reshaping the role of the CRM system and considering it as a specific back-stage on-line service for internal customers. Discussion of lessons learned from this unprecedented study concludes the paper.

Author Keywords
usability, User Experience, User-Centred Design, Services Design, collaborative design, Intranet

ACM Classification Keywords
H.1.2. Human factors; H.5.2. User interfaces; H.5.3. Group and Organization Interfaces;

INTRODUCTION
Usability of business IT systems has been a topic of numerous studies since the beginnings of HCI. Usability of company Intranets and other back-stage IT systems still has a big impact on work efficiency [1, 6, 9]. Negative experience of system operators resulting from poor usability, may also affect quality of front-line service provided for external customers.

PROBLEM DESCRIPTION

Problem background
A multi-modular CRM (Customer Relationship Management) system has been used by a large Polish financial company, but in the focus of this evaluation was included only the CRM module used by call-centre operators for serving daily hundreds of customers by the phone.

This usability evaluation project was undertaken mainly due to systematic complaints arriving from the call-centre operators, who were claiming that poor system usability dramatically slows down the–customer service. The web interface of this system was already known to be error-prone, with many operations not designed in a task-supportive manner. The system operation required plenty of very precise mouseclicks while operator's attention should be concentrated on the conversation with the customer. As a result, after reaching some critical mass, these operator complaints were seriously taken and finally a CRM usability improvement project was launched.

Evaluation framework
The financial company so far has not had their own usability staff, so an evaluation team has been formed of:

- two external usability consultants,
- four employees: the CRM system "owner" from the IT department and three very experienced senior call-centre operators.

Before starting the evaluation, following evaluation procedure was agreed with the CRM department supervisors:

1. "Crowdsourcing" will be used as a main method for gathering by e-mail all observed complaints from front-line operators in the call-centre.
2. Collaborative expert review of typical operator procedures will be performed for major operational paths.
3. Complaints collected from front-line operators will categorized by the team according to their relevance and feasibility for planned usability improvements.
4. Supplementary expert evaluation (inspection checklist and heuristic evaluation) will be applied for assessing the user interface compliance with general HCI guidelines.
5. Final report (a slideshow) will be prepared, showing prioritized recommendations and their projected impact on system usability.

Evaluation context
The team was working for several days, analyzing a live demo of on-the-phone customer service performed by senior operators. The system was operated from a laptop in a training room, with live CRM projected onto a big screen where the identified usability problems were easily visible. The demo was accompanied with narrative "user stories" by

senior operators explaining the purpose and meaning of each action performed in a call-centre conversation context.

During the presentation, operators‘ remarks and suggestions from crowdsourcing have been categorized and supplemented by senior operators’ comments on the possible impact a specific flaw could have on the customer service speed and quality.

During the teamwork we could observe gradually changing focus of attention from pure usability of the CRM system to analyzing operator’s user experience in a broader context. Moreover, in subsequent sessions it became obvious that poor usability of on-line internal services (as internal tools) affects quality of service offered to external customers.

EVALUATION RESULTS

Usability and UX aspects

Despite many usability flaws have been detected, in general in this CRM system using tab-based web interface with plenty of editable forms, operators basically met no particular problems in finding on-the-fly suitable navigation paths, matching specific needs of the actual customer on-the-phone.

However, it turned out that the most important operator UX discomforts with this CRM system were caused by some other factors, like:

- necessity to frequently quit the CRM system in order to find information available only in other systems (for instance history data of off-line contacts or access to authentication modules);
- specific technical issues, causing sudden delays in data transfer or necessity to verify currently displayed data in other systems;
- frequent new releases of the CRM system components, with no simultaneous actualization of documentation manuals etc., hence the operators had established an informal system for exchanging news about latest changes in system functionality.

The issues of demanding manual control, sub-optimal visual design or inconsistent labelling have been also raised, and later confirmed in the expert heuristic evaluation.

It also turned out that operators were very creative in finding various workarounds to overcome existing usability problems because their performance was very much affected by the bonus system, which was fed by data from automatic monitoring of operator’s actions in the CRM system. This finding was a crucial element for understanding actual operators’ work habits, motivations and attitudes, bringing important ethnographic elements to the scope of this evaluation study.

Organizational aspects

During evaluation sessions usability focus was gradually evolving towards user experience (UX) issues, interpreted in twofold manner:

(1) *Operator experience*, covering a set of emotions resulting from the CRM system behaviour and simultaneously, from the customer behaviour on the phone line, in particular:

1. demand to be polite to the customer in any circumstances,
2. demand to maintain the contact with the customer event if the CRM system is slow and there are occurring lags in data access or any other interaction problems;

(2) *Customer experience*, covering the set of emotions resulting from the perceived quality of specific on-the-phone service:

- depending on customer expectations (based on other similar services) and on actual “performance” of the service delivered;
- when the service is slow, customers often get irritated (not always expressing it in an open manner, but experienced operators feel its impact on the customer’s mood);
- repeatedly slow service on-the-phone (what sometimes appears in results of customer surveys) adversely impacts the image of the company and affects relationship with customers, who become reluctant to contact call-centre by phone.

As a result, a set of guidelines was proposed for the final evaluation report, covering issues such as:

- improvements in visual design of screens,
- software technical improvements,
- IT support quality and organizational changes,
- improving usability specifications for external software suppliers.

More importantly, a set of classified (visual, operational, performance- and feedback-related, etc.) recommendations was made, aimed at improving operators’ trust towards the CRM system and their relationship with the company brand.

Other outcomes

Apart from usability- and UX-relevant outcomes, other key findings of this study were important:

- in this project company managers for the first time decided to gather usability comments from CRM operators by open internal crowdsourcing; it produced surprisingly fruitful outcomes and

resulted in creating a unique cross-departmental cooperation around this project;

- front-line operators turned out to be highly motivated to deliver their comments in crowdsourcing and to participate in further redesign process of the CRM system; this clearly indicates a positive attitude to their work[1].

Finally, during subsequent evaluation sessions a cross-disciplinary perspective was developed in the project team, which seemed to contribute much to the project success.

CONTEXTUAL FACTORS

Success contributors

At this point, after completing the evaluation part of this project, some key success factors could be identified:

A. High commitment of staff

The first success factor - already mentioned - was very *productive crowdsourcing*, which delivered dozens of valuable comments and suggestions from the front-line.

Consequently, *senior operators and the CRM owner (IT)* - used their expertise to associate collected suggestions with specific task contexts, and were very active in searching for feasible solutions.

In both cases it was visible that the staff was aware how the usability flaws affect the service quality for external customer.

Finally, the *integrating role of senior operators* was crucial during evaluation sessions: they enabled putting the operators' complaints *into the screen context* and *into the task/organizational context*, both essential for proper adopting high-level interaction design principles to a specific screen or a conversation scene.

B. Agile-like teamwork style

The next important success factor was *agile-like evaluation cycle* which formed the canvas for the analytic part of the project. This cycle was repeated regularly for each discovered usability problem and consisted of following sequence:

1. executing step-by-step specific task situation in the CRM system, accompanied by narrative "user stories",
2. reviewing situation-relevant comments and suggestions from crowdsourcing,
3. locating and classifying user interface problems,
4. spontaneous brainstorming for possible solutions,
5. searching for the problem cause and origin,
6. problem diagnosis and reference to the procedures or local organizational context,
7. documenting proposed solution (or a set of).

This cycle was iterated for each detected problem and it allowed conducting unstructured analysis. Iterative conversational method, asking "naive" questions to the senior operators and refining answers through the unrestricted creation of ideas have finally led to developing interesting solution proposals. Moderating role of the CRM owner was very similar to the role of "scrum master" in SCRUM-based IT projects [3].

It seems noteworthy to mention that in this project creating an *ambient evaluation environment* was also very important for facilitating effective teamwork: a round table configuration, circular information flow, ongoing visual contact, a wall-size projected CRM screen as a central focus of attention - all these elements all helped to stimulate group dynamics in this project.

Novel evaluation elements

As well as direct outcomes aimed at the CRM system redesign, three methodological innovations emerged.

A. Innovation and creativity workshops

When developing proposals for improving the operator UX, both individual creativity and team-discussed refinements were combined, using spontaneous brainstorming and also analytic conceptual refinements.

Starting from visions of specific screens with improved interaction elements, the amount of creativity input was growing so fast, that it gradually converted usability evaluation sessions into a sort of innovation workshop. The list of proposed improvements and innovations was long, and they could be sorted into two groups:

- ideas relevant to UX, user interface and the CRM system, aimed at improving operator UX with the CRM system;
- ideas relevant to various organizational improvements related to the back-stage activities, like staff training, horizontal communication, coordinating the human factors issues with this CRM and other IT systems, etc.

B. Forced multipoint analysis

Due to sensitivity of this project, invited external usability experts were able to operate the CRM systems only via an authorised senior operator. It resulted in an "indirect"

[1] It is quite possible that unexpectedly high staff commitment was correlated with operators' average education level; in Poland call-centre operators often are full-time university students who take this job primarily because of flexible working time schedule. However, in this study there was no opportunity to verify possible correlation between users' commitment and their education level.

system operation, without opportunity for touching the keyboard, but with very good verbal communication instead, even more helpful in understanding the task context. Paradoxically, the apparent shortage of direct experience from "feel" of the system resulted in more extensive discussions, because domain experts (senior operators) had to explain in more detail the meaning/purpose sense of each click and each operation.

There was also observed another side effect: while it was necessary for external usability experts to understand the task context, at the same time other senior operators could to look at familiar work procedures from a viewpoint of another department. Sometimes naive questions asked at this point allowed the usability experts to learn the basics, while the rest of the team was surprised by discovering differences in their work methods and step-by-step was developing a broader view of specific part of the system.

In seems that forced restrictions in access to the system apparently facilitated developing multi-point, cross-disciplinary evaluation perspective for team members.

C. The CRM system as a back-stage on-line service

Cross-disciplinary evaluation perspective has finally led to relating the CRM system to the context of the call-centre services offered to customers.

From the external customer viewpoint everything is a service, and from the operator viewpoint everything what is provided to facilitate his/her work can be also considered a service (on-line or off-line, respectively).

As such, the CRM system actually is an back-stage on-line service for operators, who in this organization can be considered as internal customers. Analogically, the other part of the system (voice interface with an operator) is the front-stage e-service aimed at external customers.

Treating an IT system holistically as e-service (internal and external), helped to identify complementary values produced for internal and for external customers, for instance:

- service speed: both external customer and the CRM operator want to complete the service as soon as possible;
- minimizing information load: both external customer and the CRM operator want to complete the service with as little information required as possible; on the other hand, there is a significant asymmetry in access to information: the CRM operator has access to full range of data about the customer and his/her history while the customer has no specific knowledge what data are actually available and needed for a particular situation;
- positive emotional experience: both external customer and CRM operator want to avoid misunderstandings or other stress-related situations despite of lack of visual contact.

KNOWLEDGE DIFFUSION

The teamwork in this project consisted of three main parts:

1. analytic - typical evaluation, based on general HCI and usability evaluation methodologies [5],
2. creative - brainstorming and evaluating solutions, ,
3. constructive - documenting redesign recommendations, to be implemented later in another project.

In both analytic and creative parts knowledge-intensive tasks have been performed, involving cross-disciplinary knowledge diffusion among team members

For developing cross-disciplinary perspective, and common understanding of the problem, the types of knowledge transfer defined in [7, 8] were taking place, especially:

- knowledge diffusion outside the team:
  - from front-line operators to senior operators,
  - from front-line operators to IT support staff (requirements, expectations, organizational and information flow issues);
- knowledge diffusion inside the team:
  - from senior operators to the CRM owner (IT) and vice versa,
  - from external usability experts to domain experts (senior operators and the CRM owner).

All the knowledge flows helped to work common understanding of detected usability and UX problems, and to develop creative solutions.

SERVICE DESIGN PERSPECTIVE

This evaluation project has raised the significance of broader UX evaluation focus, namely treating the *interactive system as a service system*, which produces value for internal and for external customers. This was the first corporate usability consultancy project in our career, where service value chain issue came to light in a very direct manner, and elements Service Design approach [10] have been applied.

The concept of service value chain proposed in the work of Heskett et al. [4], who contended that internal service quality (incl. tools for serving customers) affects employee satisfaction and job commitment. Consequently, in this case of this CRM system the operator UX has an indirect impact on customer UX and on future relationships with the work environment as a part of the internal branding.

Adopting service value chain perspective may redefine the role of HCI in current IT projects:

- while IT these days is often merely a vehicle for launching specific on-line services (internal or external), HCI and interaction design are often expected to build UX-competitive advantage and deliver value to users (customers);
- possibly better UX results may be achieved if an interactive system is designed as a service system (IT-based), aimed to offer value for specific group of customers.

Service design perspective also rises the issue if *value co-production* [4, 10]:

- in on-line *service systems* value for customer is co-produced in part by quality of human-computer interaction, but in the other part by quality of human-socioeconomic relationships relevant to actual system usage, like convenience, cost-saving, community etc.
- in on-line service *design process* value is also co-produced by participating clients/users (Value Co-Creation), what extends the current scope of User-Centred Design and UX design closer to increasingly popular the Service Design approach.

Developing profitable on-line relationships, involves mutual sharing of values produced by specific business model.

In case of on-line service systems this perspective places current HCI design practices much closer to economics, especially if the user is a conscious consumer (external, internal) willing to consume, abut also willing to co-produce value in a specific business context relationship.

SUMMARY AND CONCLUSIONS

Table 1 presents the summary of evaluation elements applied in this CRM evaluation project. The left column contains "classical" elements – evaluation methods and tools well-known form the past studies and HCI literature.

"Novel" elements are concepts, methods, tools and techniques which newly appeared in this study and were applied in a corporate context for the first time. Many novel elements in the right column suggest that the scope of this evaluation was broader that usual, the teamwork style was agile-like, and it resulted only during the project as a result of favourable contextual factors, resulting from organizational culture of this specific company. Actually, some of these novel techniques stem form organizational or marketing research, and they appeared as novel only in the context of this usability evaluation project.

This evaluation study produced several novel outcomes, unexpected at the beginning of this project: effective use of crowdsourcing, use of narrative "user stories" ethnographically presenting operators' work habits, as well as using elements of Co-Design and Value Co-Creation, characteristic for the Service Design approach.

This project also led to a deeper understanding that:

- in e-business systems projects HCI has many common points with service design,
- many interactive systems can be designed as IT-based service systems, producing value for both internal and external customers,
- in usability evaluation and UX design users/customers should be involved as value co-producers, what extends their role in the current UCD approach.

Classical elements	Novel elements
• usability checklists • heuristic evaluation • in-depth interviews • contextual analysis • ethnographic perspective • slideshow as evaluation results reporting	• crowdsourcing • user stories • brainstorming • knowledge diffusion • high commitment of system operators • agile-like teamwork style • naive questioning • multipoint analysis • indirect system operation • Service Design perspective • understanding work behaviour: staff motivation, attitude and incentives

Table 1. Classical and novel elements applied in the CRM evaluation project

From the HCI methodology perspective two outcomes seem to be most significant:

- merging classical (from the HCI area) and novel elements, partly adopted from other studies, partly invented on the spot;
- applying *Service Design* perspective along HCI/usability research focus.

The general outcomes of this work show that in UX/usability consultancy practice there are some factors, which – if appropriately triggered – may utilise local

resources that invigorate functions driving project forward through enhanced knowledge and commitment of project stakeholders [2, 11]. In this project resources that were crucial could be identified as:

- procedural, relevant to the novel evaluation framework, integrating diverse methods in a flexible manner;
- expressive, relevant to the problem-focused group of committed CRM system operators, sketching solutions on-the-fly;
- knowledge resources, that have been identified in the project and purposefully used for a better understanding the CRM system context and values shared by its external and internal customers.

As a result, this project has proved the service value chain concept may be applied for many corporate IT systems, which should be treated as e-services designed jointly with User-Centred and Service Design approaches.

Further research work in this area is planned, because the impact of economic factors shaping behaviour of humans involved in diverse value chains/networks on-line becomes more and more significant element of user-service interactions.

ACKNOWLEDGEMENTS

This work was partly supported by the Polish National Science Centre under the contract No. 2011/01/M/HS4/04995.

REFERENCES

1. Bernard R. (1998). The Corporate Intranet. Wiley.
2. Cockton, G. (2012). Making Designing Worth Worth Designing. Position Paper, CHI 2012 Workshop on Methods for Accounting for Values in Human-Centered Computing. URL: http://ii.tudelft.nl/ValuesInDesign/submissions/cockton.pdf
3. Cohn M. (2009). Succeeding with Agile: Software Development Using Scrum. Addison-Wesley.
4. Heskett J.L., Jones T.O., Loveman G.W. (1994). Putting the Service-Profit Chain to Work. Harvard Business Review, March-April 1994. 164-174.
5. Landauer, T. K. (1988). Research Methods in Human-Computer Interaction. In: M.Helander (Ed.), Handbook of Human-Computer Interaction (pp. 905-928). Elsevier, New York.
6. Hix D., Hartson R. (1993). Developing User Interfaces: Ensuring Usability Through Product and Process. John Wiley & Sons, New York.
7. Nonaka, I. (1994). A Dynamic Theory of Organizational Knowledge Creation. Organization Science 5 (1). 14-37.
8. Probst G, Raub S., Romhardt K. (2000). Managing Knowledge: Building Blocks for Success. Wiley.
9. Shneiderman B., Plaisant C. (1994). Designing the User Interface: Strategies for Effective Human-Computer Interaction. Addison-Wesley, Reading, MA.
10. Stickdorn M., Schneider J. (2010). This is Service Design Thinking. BIS Publishers, Amsterdam.
11. 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, in IJHCI, 27(10), 940-970. URL: http://www.kasperhornbaek.dk%2Fpapers%2FIJHCI2011_WoolrychEtAl.pdf&ei=yBAeUbaOM6ag0QWvs4HYBg

“Should this be in my profile?”: Using a card sort to understand thoughts on information privacy

Edward S. De Guzman
Autodesk, Inc.
San Francisco, Calif., USA
edward.deguzman@autodesk.com

ABSTRACT

Card sorting is a technique that many information architects and user experience professionals use as an input to the structure of a site or product. In this case study we describe how card sorting has been used in my research at Autodesk in a different context: In one research study a card sort was used to elicit opinions about the privacy of information that makes up an Autodesk customer's user profile. In another research study a card sort was used as a tool to give Autodesk users a vehicle for critiquing a user interface's visual design without having a visual design background. We begin by briefly describing Autodesk and my role in the company and then describe the case study where card sorting was transferred from an information architecture context to the contexts of online user profile page interaction design.

BACKGROUND

Autodesk is a world leader in 3D design, engineering and entertainment software. Our customers span across the manufacturing, architecture, building, construction, and media and entertainment industries. Autodesk software empowers our customers to design, visualize, and simulate their ideas before they are ever built or created.

My role at Autodesk is titled: Senior User Experience Researcher. A central part of my role is to regularly meet with product managers and user experience designers. These colleagues come to me with research questions. These can range from formative research questions such as "How do Autodesk users feel about cloud computing" to more evaluative research questions such as "Please assess the usability of this product before it is released". Based I the research question I select the appropriate method and execute the research study, and report findings to the stakeholders. I have found card sorting to be a versatile method for collecting feedback in a variety of contexts.

CARD SORTING TO LEARN WHAT SHOULD BE IN THE AUTODESK USER’S ONLINE PROFILE

Card sorts are commonly used to generate a user-centered taxonomy of pages or categories for a website. This taxonomy is then used to inspire the design of a usable and intuitive navigation structure for the website. This study was a formative research study to inform the design of a profile page for a user of Autodesk software. The primary research question was: "What kinds of data are appropriate / inappropriate to include in a user profile for a web application?" The primary stakeholder was a user experience designer who wanted to use this research to inform which kinds of personal data should be included on the profile page. The stakeholder preferred that she would like at least 25 participants in this research study.

If this research study was run as a “traditional” card sort, I would have to schedule at least 25 one-on-one in-person sessions and guide each participant through the card sort activity as the moderator. Additionally, other researchers would be needed to serve as note takers and to manage the recording setup. Executing this research using traditional methods would be costly in terms of time and staffing. Since I was the only researcher assigned to this study, I needed to select an appropriate method of collecting the data in a reasonable amount of time. I used an online card sorting tool called Optimal Sort [1] to satisfy the constraint of collecting a relatively large sample size in a reasonable amount of time.

In the exercise, participants were first given instructions on how to use the online card sorting tool. Then they were asked to sort pieces of information that would appear in an online profile (for example: name, personal photo, email address) into one of three categories: "My colleagues can see this in my profile", "Only I can see this in my profile" and "This should not be in my profile".

Optimal Sort runs in a web browser (see Figure 1.) and participants were able to complete the card sort exercise on their own time but within a week of receiving the instructions from the researchers.

RESULTS

Figure 2 shows how 37 participants in this study sorted 29 types of information commonly found in an online profile. One outcome of this study is the validation of the current user profile. Most participants sorted the fields that are currently visible as "My colleagues can see this in my profile". The data also showed that an Autodesk user profile is seen more as a professional web property than a personal one. Most participants felt that a link to the user's LinkedIn profile does belong on the Autodesk User Profile but a link to a Facebook profile does not. As expected participants in the study generally agreed that some personal information does NOT belong in the Autodesk user profile, including Social Security Number, Date of

Birth, Cell phone number, and home address. Some pieces of information fell into a "gray area" where it could not be concluded whether the information is or is NOT appropriate for an Autodesk user profile. Data in this category included: University attended, degrees obtained and work history or past employment.

Triumph:

I would consider this specific case study a success. We were able to run this research study as an un-moderated online card sort exercise, using an online card sort web service. As a result, we were able to collect data from 37 Autodesk customers in five days. If we executed this study as a series of one-on-one in-person interviews. The larger sample of users resulted in a greater amount of credibility and buy-in from our stakeholders.

I acknowledge there are several potential drawbacks to conducting a card sort in an un-moderated setting. The participants might be less attentive while completing the exercise or may misunderstand the instructions. Furthermore, un-moderated studies make it difficult to collect qualitative data to understand *why* certain pieces of information were sorted into specific groups. In future rounds of research, one way to mitigate these drawbacks is to schedule follow-up interviews with a random sample of the participants. Doing so will allow us to collect a sample of qualitative data and assess whether or not the participant understood the instructions of the exercise.

ANALYSIS OF CONTEXTUAL FACTORS

This study was resourced to be executed over three to four days of a work week. Conducting 37 one-hour one-on-one interviews over three to four days is not a reasonable request for a user researcher at Autodesk. Therefore, to gather this much data, a new method needed to be employed. In-person moderated card sorting sessions may have been more feasible within the given time frame of a few days but the data collection and analysis would be slow, having only one researcher assigned to this project. Therefore, an UNMODERATED card-sort exercise nicely fit the constraints of collecting a large amount of data in a reasonable amount of time.

FUTURE WORK

Encouraged by the success of using card sorting in a context OUTSIDE of information architecture, researchers at Autodesk are exploring additional contexts where card sorting may be appropriate. One promising direction is using card sorting in research aimed at understanding users' initial impressions about visual design. Several pilot studies have been run where the participant is shown a high-fidelity mockup or screenshot as a stimulus. Then the participant is given a set of cards. On each card is a single word or phrase that would come up in a conversation about visual design. For example there are cards for: Clean, Simplistic, Fresh, Overwhelming, Friendly, Unattractive. The participant is asked "How does this look?" and must respond by selecting three to five cards that match how the design appears to them. This could become another context of method transfer -- in this case, card sorting moves from an information architecture context to a visual design critique context. More work is needed to refine the Visual Design Card Sort technique and understand when it is appropriate to apply this research method.

CONCLUSION

In this case study we describe an example of method transfer: card sorting, a technique traditionally employed in information architecture is successfully used to understand user opinions about information privacy. Because of the expectation for a large set of data to be collected in a short amount of time, traditional interview protocols were not used. Instead an unmoderated card sort activity was used to collect the data. As a successful approach in this case study, Autodesk is exploring other ways to use unmoderated card sorting as a research technique. We are currently exploring how a card sort can be used to help users give critical feedback on visual design without having to be visual design experts.

REFERENCES

[1] Optimal Sort
http://www.optimalworkshop.com/optimalsort.htm

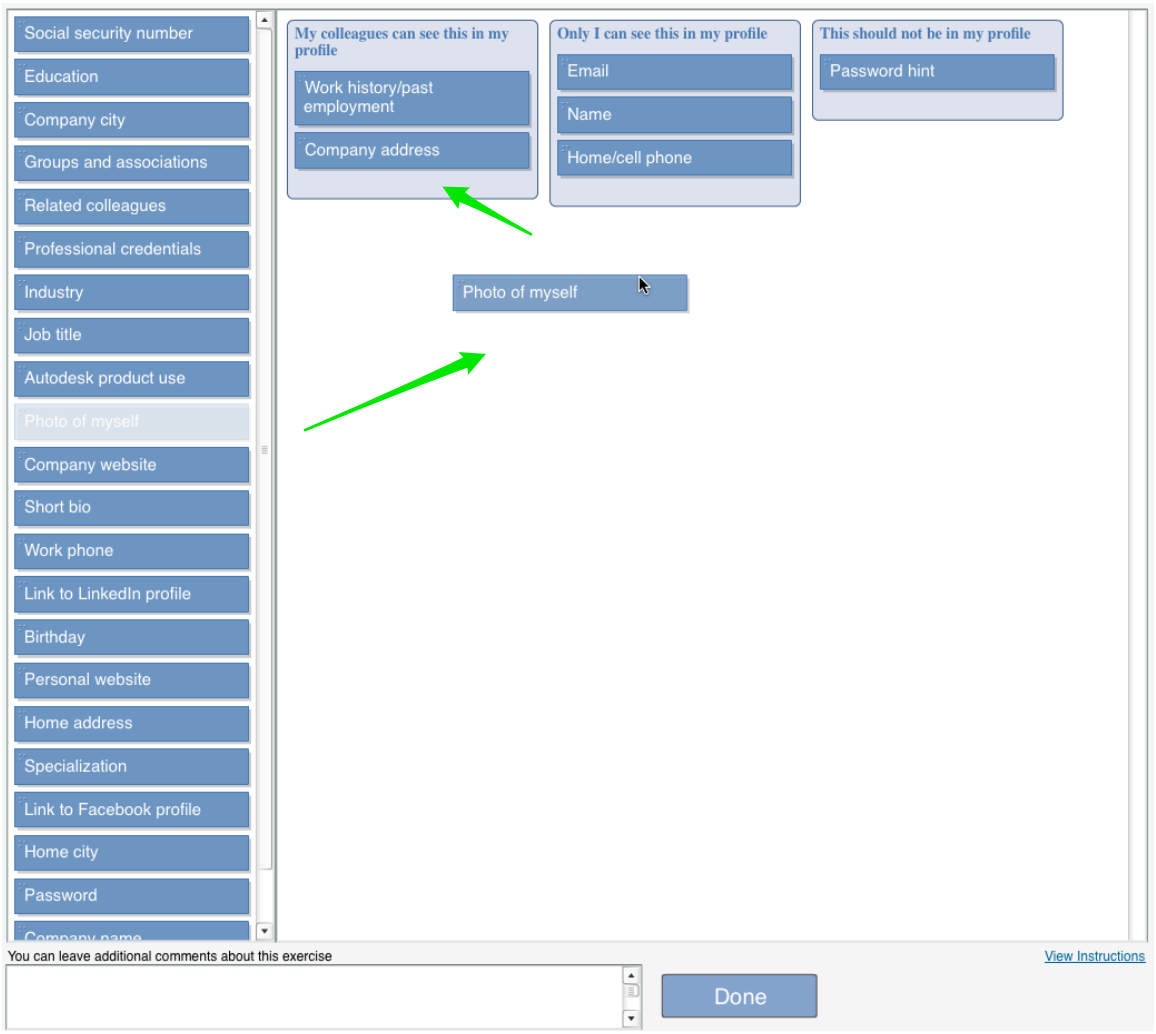


Figure 1. Screen shot of Optimal Sort interface

	My colleagues can see this in my profile	Only I can see this in my profile	This should not be in my profile
Industry	100%	0%	0%
Company name	97%	3%	0%
Job title	95%	5%	0%
Name	95%	5%	0%
Company city	89%	5%	5%
Company website	89%	8%	3%
Specialization	86%	3%	11%
Photo of myself	84%	5%	11%
Autodesk product use	76%	19%	5%
Groups and associations	76%	0%	24%
Professional credentials	76%	5%	19%
Short bio	76%	5%	19%
Link to LinkedIn profile	73%	3%	24%
Email	70%	24%	5%
Company address	65%	24%	11%
Work phone	62%	32%	5%
Personal website	59%	3%	38%
Related colleagues	59%	27%	14%
Password hint	0%	81%	19%
Password	0%	57%	43%
Social security number	0%	5%	95%
Birthday	14%	22%	65%
Home/cell phone	8%	30%	62%
Home address	3%	41%	57%
Link to Facebook profile	35%	11%	54%
Education: school & degree	41%	14%	46%
Work history/past employme	27%	30%	43%
Education	46%	16%	38%
Home city	41%	27%	32%

Figure 2. Results from Card Sort Exercise

Heuristic Evaluation of User Experience – Case Nokia

Heli Rantavuo
Nokia Design
2 Kingdom Street, Sheldon Square, W2 6BD, London
heli.rantavuo@nokia.com

Virpi Roto
Aalto University
School of Arts, Design, and Architecture
P.O.Box 31000, 00076 Aalto, Finland
virpi.roto@aalto.fi

ABSTRACT
Along several years, Nokia has utilized user experience heuristics for assessing design prototypes with a panel of internal experts. In this paper, we describe how we have tailored the traditional heuristic evaluation method to suit the needs of a large multinational corporate.

HEURISTIC EVALUATION
Heuristic evaluation method [3] has traditionally been used for evaluating usability against a set of principles, e.g. the 10 usability heuristics by Jakob Nielsen [4]. Heuristic evaluation of user experience (UX) in a broader sense seems to be rare, which, we believe, is due to the situational nature of UX. It is hard to derive universal UX heuristics that would imply good UX for all kinds of users in various different situations with various kinds of products and services. This is a core reason why UX heuristics need to be tailored for the different evaluation cases. The few examples of UX heuristics have indeed been developed for specific cases: UX of cross-platform social networking systems [7], playful experiences of games [1], and persuasive health technologies [2].

A few years ago, we reported heuristic evaluation as one of the most potential methods in the early phase of product development [6]. The main benefits include

- fast application of the method after building the infrastructure for systematic heuristic evaluation
- possibility to run the evaluation in different parts of the world simultaneously
- possibility to evaluate early ideas that are not articulated very clearly yet
- experts can understand the future developments that will influence the UX of futuristic concepts.

The main challenges in heuristic evaluation of UX include setting up the infrastructure for systematic and rapid heuristic evaluation, and foreseeing how ordinary users will experience the concept once it comes out. The heuristics should be formulated so that the latter challenge can be alleviated.

OVERVIEW OF THE EVALUATION CONTEXT
UX123 is an infrastructure established inside Nokia for evaluating the UX design competitiveness of mobile products in a systematic manner. The UX123 team serves product development with a pool of internal experts on different sites of the company. A set of UX heuristics has been developed to make the evaluation systematic. Whenever a new evaluation is needed, a subset of evaluators is recruited to assess the UX of a given user journey on the given products against the heuristics. Competitor products are assessed with the same set of heuristics. The evaluators report their findings in a review workshop and the UX123 team analyses and categorizes the findings and reports to the product team as well as other stakeholders. The evaluation can be done in different phases of product development.

Below we describe how the method-resources [8] have been tailored over time in different evaluation cases within Nokia.

Participant recruitment
The principles of expert recruitment are the same in all evaluation cases: experts are senior UX design professionals with experience from product development both inside and outside of Nokia.

Design representation
Since heuristic evaluation is conducted in various different phases of product development, the representation of the design that experts evaluate varies accordingly. In the early phase, the representation may be a concept demonstration, in a later phase an actual prototype. Experience has shown that UX heuristics are best applied to concepts that allow or simulate interaction. Concepts without interactive aspects tend to demonstrate aspects such as visual design well, whereas other aspects such as input mechanisms often cannot be assessed. As a result, all of the heuristics may not be applicable and a holistic assessment of the UX may not be possible.

The UX123 evaluation process includes competitor evaluations. Naturally, these devices are products on the market.

Heuristics
As described in Introduction, the UX heuristics need to be domain specific. UX123 team has developed a set of UX heuristics that address, and are broken down from three main categories: the aspect of pleasure or delight when

using mobile devices and services; smooth task flow; and actual and perceived effort. The heuristics reflect both current understanding of what enables good UX in this domain and Nokia's strategic aims in increasing its competitiveness. The heuristics are the same for all UX123 evaluation cases, but the significance of the different sections within the heuristics varies, according to what is the competitive situation of the evaluated product. For example, in some cases there may be specific interest in how competitive the physical design of a product is, whereas in another review it may be more significant how competitive is the social networking experience.

Task definition

Since the products are different in the different evaluation cases, also the tasks that the experts do during the evaluation differ case by case. As user experience is highly situated, the tasks will be tied to a user story that describes a potential usage situation. Without imagining the use situation, it is hard for experts to project the end user experience. The product or service development team together with the UX123 team defines the suitable set of stories for the evaluation.

Problem identification

As we outlined in the Introduction, UX123 evaluates the competitiveness of UX design in mobile products. Problems are identified from two perspectives. First, from the perspective of how good the UX of the product is overall. Second, from the perspective of the specific competitive situation of the product. A problem found from the perspective of good UX design becomes less significant from the competitive angle if it is shared by all competitors. On the other hand, something that may not negatively impact the existing UX, for example the absence of a feature, may become a problem from the competitive angle if competitors offer that feature and it becomes a user expectation.

Problem classification

According to the guidance of heuristic evaluation, the experts classify the found problems by indicating the related heuristic for each problem. They classify the problems under the heuristics. This was also the case in UX123 when the experts were not very familiar with the heuristics.

Once the experts got familiar with the heuristics, they did not check the list in detail anymore. They rather reported found problems either under the main theme of the heuristics or without explicit reference to specific heuristics at all. The responsibility of precise problem classification thus shifted from experts to the UX123 team that collected and analyzed the experts' reports. The benefit of this is that when assessing the products, experts can focus on identifying and describing problems and not worry about classification, while still conscious of the specific principles and focus outlined by the UX123 heuristics.

When it comes to assessing problem severity and fixing, UX123 provides a suggestion based on competitive analysis for the product development team, but the team and product owners ultimately decide which problems to prioritize when improvements are made. In other words, the responsibility for fixes or changes in the product lies not with the UX123, the UX assessment program, but with the product team and the relevant stakeholders. When it is relevant and possible, the UX123 team follows up the fixes and improvements to the problems pointed out in reviews, in order to track the impact of the assessment work.

Reporting format

In the early days of UX123, the evaluation results were used to raise awareness on UX and flag the importance of competitive UX design within the company. This meant the reports were optimized for a strong impact at the management level, through rich visual material that illustrated problems as well as strengths found in UX competitiveness, and through comprehensive argumentation that would be hard to dismiss without action. This technique of influencing was successful, and UX123 gained visibility, trust and an influential position in the organization.

Thanks to this initial impact, UX123 became an established activity supporting the product development teams as well as senior leaders in decision-making. The reports now serve the teams and leaders by providing data from the heuristic evaluation. Reporting has shifted from heavy reports with thorough visual evidence to a combination of spreadsheets describing problems and strengths in detail, relevant for product development, and executive summaries that show more competitive analysis, relevant to leadership. Reporting formats are adjusted according to the needs in each evaluation case.

HOW COMPANY STRATEGY AFFECTS THE METHOD

Companies investing in UX naturally do so to make their customers happier, but also to differentiate from competition and strengthen the brand image [5]. Designing products and services that would enable UX unique to a specific company requires in-depth investigation of the brand and mission of the company, as well as those of the competitors. We see it essential that any wider heuristic UX evaluation initiative is aware of and adjusts its goals and operations to the company strategy.

When UX design is not present in the company strategy as a factor in gaining competitiveness, along with the task of the UX evaluation itself it becomes an important task to exert influence and argue for the significance of the findings. The UX123 program was created in this type of a context and was successful in the task of influencing through rigorous competitive analysis and thorough reporting that provided concrete examples of competitive strengths and weaknesses. In global scale product and service development no product or business decision is

made based on one data source only. UX123, however, became one of the significant sources for the momentum where the end user perspective and UX gained more ground in decision making.

In cases where UX design is present in the company's strategy and is recognised as an asset for competitiveness, which is the current context for the UX123 program, it becomes purposeful for a heuristic UX evaluation program to draw from the strategy. It may be possible to derive specific heuristics from the strategic goals, and when reporting the results to decision makers and aiming for an impact, references to the strategy are effective.

CONCLUSION

If we had applied heuristic evaluation as it was defined in 1994 [4], the attempt would have most likely been a failure: the impact on making more competitive products would have been extremely limited, and the exercise would have been short lived. By adjusting the method to cover UX aspects and comparison to competitors, by drawing the expertise from a wide network of internal professionals, by the core team taking responsibility for problem classification and by finding suitable reporting formats, the UX123 program in Nokia has been highly successful. The evaluations have had direct impact on products and services, either resulting in design changes or supporting designs that enable good UX but may increase cost or development time. Indirectly, across designer and decision maker levels, the program and the evaluations have increased awareness of current UX competitiveness, and UX design as a competitive asset in general. This awareness ultimately has an impact in decision making around entire product portfolios and company strategy.

REFERENCES

[1] Desurvire, H., Caplan, M. and Toth, J.A. 2004. Using heuristics to evaluate the playability of games. In *CHI '04 Extended Abstracts on Human Factors in Computing Systems* (CHI EA '04). ACM, New York, NY, USA, 1509-1512.

[2] Kientz, J.A., Choe, E.K., Birch, B., Maharaj, R., Fonville, A., Glasson, C., and Mundt, J. 2010. Heuristic evaluation of persuasive health technologies. In *Proceedings of the 1st ACM International Health Informatics Symposium* (IHI '10), Tiffany Veinot (Ed.). ACM, New York, NY, USA, 555-564.

[3] Nielsen, J. and Molich. R. 1990. Heuristic evaluation of user interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '90). ACM, New York, NY, USA, 249-256.

[4] Nielsen, J. (1994). Heuristic evaluation. In *Nielsen, J., and Mack, R.L. (Eds.), Usability Inspection Methods*, John Wiley & Sons, New York, NY

[5] Pine, B. J., & Gilmore, J. H. (1998). Welcome to the experience economy. *Harvard business review*, *76*, 97-105.

[6] Roto, V., Rantavuo, H., Väänänen-Vainio-Mattila, K. 2009. Evaluating User Experience of Early Product Concepts. In *Proceedings of Designing Pleasurable Products and Interfaces* (DPPI'09). Compiegne, France.

[7] Väänänen-Vainio-Mattila, K. and Wäljas, M. 2009. Developing an expert evaluation method for user eXperience of cross-platform web services. In *Proceedings of the 13th International MindTrek Conference: Everyday Life in the Ubiquitous Era* (MindTrek '09). ACM, New York, NY, USA, 162-169.

[8] Woolrych, A., Hornbæk, K., Frøkjær, E., & Cockton, G. (2011). Ingredients and meals rather than recipes: A proposal for research that does not treat usability evaluation methods as indivisible wholes. International Journal of Human Computer Interaction, 27(10), 940-70

Facilitating the take-up of new HCI practices: a 'diffusion of innovations' perspective

Arnold Vermeeren
Industrial Design Engineering,
Delft University of Technology
Landbergstraat 15
2628 CE Delft, The Netherlands
a.p.o.s.vermeeren@tudelft.nl

Gilbert Cockton
Department of Media and
Communication Design
Northumbria University
Newcastle upon Tyne, NE1 8ST, UK
gilbert.cockton@northumbria.ac.uk

ABSTRACT
The workshop *Made for Sharing: HCI Stories of Transfer, Triumph & Tragedy* focuses on collecting cases in which practitioners have used their HCI methods in new contexts. For analyzing the collected body of cases we propose to apply a framework inspired by the *Diffusion of Innovations* approach which focuses on what facilitates the adoption, re-invention and implementation of new practices in social systems.

Author Keywords
Human-computer interaction; user experience; usability; diffusion of innovation; methods.

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

INTRODUCTION
The workshop *Made for Sharing: HCI Stories of Transfer, Triumph & Tragedy* [9] focuses on "Understanding, via structured case studies, how HCI professionals transfer the same (set of) design and evaluation methods across use contexts in terms of appropriating and configuring method-resources". Based on his empirical and theoretical work on adoption and adaptation of usability evaluation methods, Furniss [7] stressed that "adoption and adaptation cannot be fully understood devoid of context". Therefore, this workshop "intends to generate insights in the design work required to get HCI methods to work, and how this is impacted by contextual factors such as application domains, organizational factors and project constraints." In this position paper we propose an initial framework for structuring the findings from the case studies, that is inspired by Rogers' work on *Diffusion of Innovations* [12].

Transfer of HCI methods seen as Diffusion of Innovation
Rogers defines *an innovation* as "an idea, practice, or object that is perceived as new by an individual or other unit of adoption." Diffusion is defined "as the process by which (1) an *innovation* (2) is *communicated* through certain *channels* (3) over *time* (4) among the members of a *social system*." In this paper we treat HCI methods or approaches that are applied in a new context as innovations.

Applying the diffusion of innovations framework means that the new context is seen from the perspective of a *social system*, and that next to social context factors, *communication* is seen as playing an important role in the adoption and implementation of new methods. Furthermore, the diffusion of innovations approach implies a *process view* of adoption, adaptation and implementation, rather than a static view on matching characteristics of a context to attributes of a method.

Methods and Innovations are no Indivisible Wholes
According to Rogers [12] "Until about the mid-1970s, it was assumed that an innovation was an invariant quality that was not changed as it diffused" and since then some researchers started seeing *re-invention* as "the degree to which an individual's use of a new idea departed from the "mainline" version of the innovation" or as "the degree to which an innovation is changed or modified by a user in the process of its adoption and implementation". This view is very much in line with Woolrych et al's view on not seeing HCI methods as indivisible wholes [13]. Rogers [12] emphasizes that "We should remember, therefore, that […] adopting an innovation is not necessarily a passive role of just implementing a standard template of the new idea." The present workshop focuses on exactly this process of changing or modifying HCI methods when implementing them in a new context.

METHOD TRANSFER AS DIFFUSION FROM A SOCIAL COMMUNICATIONS PERSPECTIVE
HCI method selection tools (e.g., UsabilityPlanner.org, UCDtoolbox.com, AllaboutUX.org) provide their users with assistance in finding appropriate methods or approaches for specific contexts. They do this largely based on matching method attributes to (presumed) attributes of the target context in which they will be used. Such tools may also provide advice on how to adapt methods to the situations. This was also the original approach taken in the EU Cost Action IC0904 TwinTide project [http://TwinTide.org], as well as (implicitly) in the EU COST action 294 MAUSE [http://cost294.org]. More recently, there has been a shift in the TwinTide project towards a focus on the *process* of transfer. The framework we propose in this paper attempts to connect the two approaches, via *social system* and *communication* perspectives.

Part 1: Method Transfer as Resource Matching

As already expressed in the Introduction section, when discussing the use of methods in new contexts we don't see methods as indivisible wholes. We even consider the word 'method' as misleading in the sense that it suggests that a method is an invariant entity, a fixed set of procedures, materials, etc. For example, this could refer to its 'mainline' version (cf. Rogers [12] for innovations) or the version as it was originally (intended to be) applied by its developers. Following Woolrych et al. [13] we would rather conceptualize *method attributes* as *resources* for *approaches* to HCI work. Matching attributes then comes down to matching (approach groups of) resources to the work to be done and the values, skills, experiences (etc.) of the people that need to do that work. Approaches always are incomplete and always require modification, replacement or addition of resources for specific design work needs.

One way to expose gaps in an approach's resources is via Woolrych et al's [13] categorization of resource types for usability evaluation work. In the TwinTide project we have taken a broader view, not only considering usability evaluation, but HCI design and evaluation work. Based on various workshops we have so far come up with an adapted categorization of resources [5], and in TwinTide we now see resources as having functions, rather than types (more accurately, resource have several attributes, of which one is the types of functions that they perform). This is because 'types' suggests a simple resource taxonomy, whereas 'functions' suggests something about the action or activity for which it is used, and are only one attribute of a resource (the form of its materialization is another). For many resources it is problematic to categorize them as being of a certain type, as they can have more than one function (i.e., have multiple practical ways of using them). For example, as well as communicating ideas, sketching can also support their generation and structure the process of selecting and refining promising options.

Scoping and axiological functions

According to Cockton [5] scoping and axiological resources, express the intended coverage, motivating values and proscribed practices of approaches. *Scoping resource functions* indicate the extent of a method's applicability in terms of the purposes and usage contexts of what is being designed or evaluated, including application areas/domains [13]. These functions relate to issues such as the extent to which approaches are intended for specific target groups, specific application areas, or for specific activities (e.g., analysis, rather than creation). Approaches can also be scoped by technology (e.g., ambient display heuristics [10]) or application domain (e.g., games [6]). Further to this, approaches may be focused on specific design choices, such as choosing how users should benefit, or choosing user interface features.

Scoping resources support rapid initial matching by design teams looking for new approaches through their focus on development phases, target users, application domains or sectors, or technologies. However, an approach that may appear too general (e.g., games heuristics) may be modifiable for a specific genre (e.g., sport games). Similarly, specialized heuristics (e.g., for ambient displays) may transfer to loosely related technologies (e.g., splash screens on kiosks or in games).

Axiological resource functions indicate the values underpinning a method (perspectives) [13] (axiology is the study of values). These resources relate to for example ethical considerations in using certain approaches, as well as to factors related to the disciplines from which an approach originates (as disciplines bring along specific systems of values, content and method). For example, an evaluation from a psychological perspective is based on a different axiology than an evaluation from a software engineering or sociological perspective. Discount methods value cost reductions, which would be appealing in design contexts where budget is not available for extensive user experience work. In contexts where user-centered design is highly valued for its benefits, discount methods may be less attractive.

Axiological resources support rapid initial matching by design teams looking for new approaches in a similar manner to scoping resources. Teams needing to minimize costs will be drawn to discounting values, while teams developing high integrity systems will be drawn to approaches that prioritize valid results.

Harvesting functions

Woolrych [13] defines instrumentation resources as 'resources to collect issues and measures for evaluations'. Cockton [5] broadens this to harvesting resources to also include creative design activities rather than evaluations only. A resource has a harvesting function when it collects data, both for contextual research and evaluation, but also for design inspirations and directions. In evaluations, examples of harvesting resources are the type of data that are collected (e.g., quantitative data from surveys, eye movements, etc.) and the equipment needed for that. In terms of inspiring designers in their contextual research or creative acts one can think of for example the use of cultural probes [8] or the materials used and the type of data one gets from participatory design activities [1].

Harvesting resources support transfer by drawing attention to potentially new information and inspiration that could fill known gaps (or previously unrecognized ones) in the inputs to design processes.

Directive functions

Cockton [5] sees directive resources as a combination of Woolrych et al's [13] procedural and project management (process) resources. Woolrych et al define procedural resources as guiding the use of a method, including partial

automation through tools. Project management (process) resources situate a method within an embracing development and collaboration context. This is now seen as a function of resources that scope approaches for particular phases of a particular design process structure (i.e., scoping function).

Directive resources are here defined as any resources that guide behavior, i.e., they direct interaction design work. Examples are the procedures that an approach prescribes. In user testing it can, e.g., refer to constraints such as an observer not being allowed to interfere with a participant's actions. In brainstorming it can refer to rules such as not criticizing ideas prematurely. Different approaches may vary in the level of formality of such procedures, in the number of prescribed procedures or the level of strictness of applying them, e.g., there are not many formal and detailed prescriptions for conducting a heuristic evaluation [11], whereas there are very detailed instructions for how to do Key Stroke Level Modeling [2]. Using procedures may also be supported by automated tools, such as SPSS for statistical analyses.

Directive resources support transfer by indicating how approaches are used in practice. Transfer will often depend on the costs of using an approach. Directive resources can indicate the work required to get an approach to work.

Expressive functions

All resources have knowledge and expressive functions (as each resource must express itself in some way, and must have a set of underlying concepts and/or knowledge). For usability evaluations Woolrych et al. [13] defined expressive resources as "communicating the output of a method via specifications, reports etc." In design, expressive resources will be chosen in relation to what a designer is trying to create or envision, e.g., for developing the aesthetics of a web site, a designer will use different expressive resources (e.g., broad nib markers or Adobe Illustrator), than for designing the navigational structure or interactivity of a website (e.g., scripting in Adobe Flash, or MS-PowerPoint). Hence, we broaden the definition of expressive resources to resources that communicate output of the use of a method or content while using it, as well as intermediate results of design work. Some expressive resources are local to designers, but others serve as boundary objects between designers and other project stakeholders.

Expressive resources support approach transfer by offering new ways for design teams to track their design work internally, as well as new forms for external communication. As with all resources, this will offer solutions to a known need, or highlight opportunities that design teams were not aware of.

Knowledge functions

In case of resources with a knowledge function, the knowledge expressed can be conceptual, theoretical or substantive, e.g., information about an approach's origin, or about its fundamental concepts (e.g., goal, task, severity). These are typically issues that are in focus about methods, in scientific or professional articles, manuals, tutorial sessions, etc.

Knowledge resources support transfer through a range of valuable benefits, including inspiration, guidance, confidence, more efficient work through re-use, and more effective design work through new capabilities. Again, these either offer to meet known needs or suggest new opportunities.

Current Developments on Resource Functions

In [5], the Working to Choose (W2C) framework integrated extensions to resource types from [13] with *Meta-Principles for Designing* [3] and *Abstract Design Situations* [4].This related scoping functions to the different types of choice and their coordination (which result in different Abstract Design Situations). Resource functions were shown to realize meta-principles.

Currently, new resource functions are being identified [14], via the distinction between *expressive* functions (local to designers) and *performative* functions (communication with design stakeholders), identification of *emotional* functions for some design resources (through propelling or caring for the design process), and *integrative* functions (corresponding to meta-principles associated with co-ordination of design choices). These new functions support transfer by offering improved internal audit trails or external communication, more dynamic and less frustrating work cultures, and more effective integration of design inputs, activities and results.

Part 2: Diffusion of approaches in practice

Part 1 highlighted resource functions that play a role in determining if an approach can be used in a specific context and what needs to be modified or added to how an approach has been implemented in a preceding context. Part of the process of implementing an approach in a new situation is trying to match the various resources as objectively as possible, in order to find an appropriate fit for the work to be done. Insights from diffusion of innovation research add a further perspective. This perspective makes clear that even if there seems to be a perfect fit, there are other factors that play a role in deciding on an approach or on how to implement an approach. This perspective relates to the social context in which practitioner work. Below we will discuss three groups of findings from diffusion of research that seem relevant to our case: *adoption-relevant attributes of innovations*, *change agent success factors*, and the *innovation decision process*.

Adoption-relevant attributes of innovations

Rogers [12] mentions five main attributes of innovations that play a role in whether an innovation will be adopted in a social context or not. In our cases, when practitioners start working in a new (social) context, wanting to apply an

approach they are familiar with, a similar situation may occur: not only should the practitioners themselves find a match between the approach and the work to be done, they will also be confronted with some social context in which they work. To this social context, the new approach may be an innovation, and usually they may have to modify an approach to increase the chance of a successful implementation. According to Rogers [12] the following five attributes of an innovation (here: approach) as perceived by the members of a social system may play a role in the adoption process: 1) *relative advantage*, 2) *compatibility*, 3) *complexity*, 4) *trialability* and 5) *observability*.

Relative advantage is the degree to which an innovation is perceived as being better than the idea it supersedes [12]. When practitioners introduce a new approach in their social environment, it has a better chance of being adopted if the people in that social environment *perceive* the approach as having a relative advantage. Note that the word *perceive* is as crucial as *relative advantage* here (as it is with the following four attributes). If the practictioner's environment doesn't see the relative advantage there is a higher chance that they will resist the change in their usual way of doing things. *Compatibility* is the degree to which an innovation is perceived as consistent with the existing values, past experiences, and needs of potential adopters. For HCI approaches as innovations, this is largely related to the match of resources discussed in part 1, however, this attribute emphasizes that what matters, is how social contexts shape perceptions of matches. *Complexity* is the degree to which an innovation is perceived as relatively difficult to understand and to use. If the social context thinks a new approach is difficult to use, or if they don't understand it, this lessens the chance of them agreeing about using it. *Trialibility* is the degree to which an innovation may be experimented with on a limited basis. So if an approach can be tried out on a limited scale without too many risks, this helps in introducing it. *Observability* is the degree to which the results of an innovation are visible to others. The results of some ideas are easily observed and communicated to others, whereas some innovations are difficult to describe to others. The same is true for new approaches. If after using a new approach it is difficult to observe or describe whether there is any difference in results or not, this lessens the chance of an approach being adopted.

Change agent success factors

Much of the diffusion of innovations research is about the role of change agents. About change agents Rogers states: "A change agent is an individual who influences clients' innovation decisions in a direction deemed desirable by a change agency. In most cases a change agent seeks to secure the adoption of new ideas, but he or she may also attempt to slow the diffusion process and prevent the adoption of certain innovations." Rogers concludes that "a change agent's relative success in securing adoption of innovations is positively related to 8 factors: (1) the extent of change agent effort in contacting clients, (2) a client-orientation, rather than a change agency-orientation, (3) the degree to which the diffusion program is compatible with clients' needs, (4) the change agent's empathy with clients, (5) his or her homophily with clients (homophily is compatibility as the degree to which pairs of individuals who interact are similar in certain attributes, such as beliefs, education, social status, and the like [12]), (6) credibility in the clients' eyes, (7) the extent to which he or she works through opinion leaders, and (8) increasing clients' ability to evaluate innovations." Considering the situation of a practitioner entering a new context some of these may be considered relevant as well. Effort in contacting the client (1) doesn't seem to be relevant here, as we assume that the practitioner is in the same team. This would also mean that client-orientation and change agent's orientation (2) will generally be the same. Furthermore, the situations we consider do not deal with diffusion programs (3) deliberately aimed at spreading certain practices just for the sake of spreading them. Increasing a client's ability to evaluate innovations (8) comes down to change agents seeking to raise the clients' technical competence and ability to evaluate potential innovations themselves. This is a long-range endeavor, which is also not relevant to the cases we consider here. What remains are empathy (4), homophily (5), credibility (6) and opinion leaders (7). For HCI practitioners wanting to introduce new approaches into a new context, this means that this will be more easy if the practitioner shows *empathy with other team members*, *is more homophilous with them*, if other team members *see the practitioner as credible*, and if the change agent can refer to *other teams or people that use the practitioner's approach* and who are seen by the team as *opinion leaders*.

The innovation decision process

Rogers [12] defines the innovation-decision process as "the process through which an individual (or other decision-making unit) passes from first knowledge of an innovation, to forming an attitude toward the innovation, to a decision to adopt or reject, to implementation of the new idea, and to confirmation of this decision." Rogers originally distinguished five stages in the innovation decision process [12]. Although he consistently talks about an individual *or other decision making unit*, these stages seem to relate to individuals making choices mostly. These stages are 1) the (awareness-) *knowledge* stage when the individual (or other decision making unit) is exposed to the innovation's existence and gains some understanding of how it functions, 2) the *persuasion* stage in which one may become interested in the innovation and starts forming a favorable or unfavorable attitude towards it, 3) the *decision* stage when activities are undertaken that lead to adopting or rejecting the innovation, 4) the *implementation* stage in which an innovation is put into use, and 5) the *confirmation* stage when an individual (or other decision making unit) seeks reinforcement of an innovation-decision already

made, but he or she may reverse this previous decision if exposed to conflicting messages about the innovation. Evidence for a very clear distinction between implementation and confirmation stage is weak according to Rogers [12]. Rogers also discusses the innovation process in organizations and in that he distinguishes the following stages: 1) *agenda setting stage* in which an organization becomes aware of a problem in the organization that needs to be solved or is confronted with an innovation that uncovers a thus far unknown need, 2) *matching stage*, in which an organization is trying to figure out whether it seems worthwhile to adopt the innovation or not and tries to imagine the consequences of the innovation when implemented in the organization, 3) *redefining/ restructuring stage*, in which a solution is sought for an imperfect match between innovation and organization, either by re-inventing the innovation or by restructuring the organization, and 4) the *routinizing* stage in which the innovation becomes part of the daily life.

Transfer is thus prepared for at the agenda setting stage and then achieved via the others. If we translate this to the case of the practitioner wanting to introduce an approach to and in a new context we could summarize the process as follows:

The practitioner in the new context makes the others aware of a candidate approach or of an organizational need (knowledge stage), and makes the organization aware that a certain approach could fit an organizational need (agenda setting). To be able to apply the new approach, the practitioner needs to persuade those most directly involved in applying the approach or at least get them interested to cooperate (persuasion). For the practitioner's work to be done, he or she would need to evaluate the match of approach resources to the new situation (decision stage) and for the organization it would mean matching how it would fit the organization: what is the effect on the organization, how does it benefit the organization (matching stage and decision stage). Once the decision is taken to start using the approach the implementation phase starts, involving actually redefining the method by selecting appropriate resources and at the same time restructuring the organization (redefining/restructuring stage). Once taken into use the routinizing stage and/or confirmation stage can start.

FRAMEWORK FOR ANALYZING THE CASE STUDIES

By using this framework for analyzing cases of problems in using approaches in new contexts, one may gain a better understanding of why this happens, and learn where to search for solutions. Below, the items discussed so far are summarized and presented in a form that can be used for analyzing cases of using approaches in new contexts.

Resource functions and innovation process stages

In case of rejection of a (proposed) method, or of having to adapt it, this may occur at different stages of individual (I) or organizational (O) innovation decision processes:

- *Knowledge/awareness* (I1 - i.e., people in the new context not being aware of the method or not knowing what it can do);
- *Persuasion/interest* (I2 - i.e., difficult to get people in the new context interested or to make them form a favorable opinion about the method);
- *Agenda setting* (O1 - i.e., difficult to convince people that using the method leads to fulfilling organizational needs).

In the above stages an initial match of resources is usually being made for axiological or scoping resources. Possible reasons for not adopting an approach at these stages are:

- *Scoping:* the method does not fit the purpose of the work or the usage/process context well enough;
- *Axiological:* the method takes a different perspective on the work than is desired in the new context (e.g., with respect to what is valued), or there are ethical problems in using the method.

Reasons for not using an approach or for having to adapt it can also be found in the following stages, in which actual decisions are being made and implemented:

- *Decision/matching* (I3/O2 - i.e., difficulties in the actual process of taking the decision on whether to start using a method or not; evaluating its pros and cons; thinking through the consequences of implementation);
- *Implementation/restructuring/redefining* (I4/O3 - i.e., identified mismatches between resources and work context that lead to adaptations or modifications of the method's resources; or to changes in the organization to make it work);
- *Confirmation/routinizing* (I5/O4 - i.e., problems in sustaining a method's use).

In the above stages, considerations concerning the following resource functions play a major role. This is especially so in the decision and matching stage. However, in the later stage they continue to play a role:

- *Harvesting:* the instrumentation or type of data that the method works with, does not provide the kind of data or insights that the new context (wants to) work with;
- *Directive:* there is something about the procedures in using the method that does not fit the new context, or the procedures are perceived as being too complex or as having a poor cost-benefit ratio;
- *Expressive:* the kind of output the method gives or the way important elements are expressed with the method does not match the expectations and/or standards for communication in the new context.

Attributes of innovations

Rejecting a (proposed) method or having to adapt it often relates to attributes that are typical for innovations in general:

- *Relative advantage:* not enough relative advantage, or relative advantage not being clear enough;
- *Compatibility:* perceived problems of applying the 'old' method in the new context (i.e., team perceives a mismatch between resources and work to be done);
- *Complexity:* method perceived as being too complex to use, or too difficult to learn;
- *Trialibility:* method cannot be tried before deciding to use it;
- *Observability:* merits of the method are difficult to observe by people not directly involved in using it.

Personal (Change Agent) Factors

Sometimes application of a specific method also largely depends on personal relationships. The following change agent factors can obstruct success for someone wanting to introduce a change (e.g., a new approach):

- *Empathy:* not enough empathy between the practitioner and the new team.
- *Homophily:* difficulties in identifying with and associating with the people involved in using the new method in the new context, making them feel they are on different wave lengths.
- *Credibility:* the other people in the new team just didn't believe enough of the presented benefits of using the method.
- *Opinion leaders:* there was a lack of opinion leaders (in the eyes of the other people involved) who are also in favor of using this method**.**

CONCLUSION

A framework has been presented for analyzing cases of introducing or adopting HCI practices in new contexts. The framework combines the innovation decision process stages from Rogers [12] diffusion of innovations approach with W2C's [5] resource functions approach. Additionally, general attributes of innovations and personal factors that play a role in successful diffusion of innovations are part of the framework. Thus this framework takes a step beyond the approach most current method selection tools take, by taking social system and communication factors into account.

REFERENCES

1. Bødker, S., Ehn, P., Kammersgaard, J., Kyng, M., Sundblad, Y. A Utopian experience. In G. Bjerknes, P. Ehn, & M. Kyng. (Eds.), *Computers and democracy: A Scandinavian challenge*. Aldershot, UK: Avebury (1987), pp. 251–278.
2. Card S.K. , Moran T.P., Newell A. The keystroke-level model for user performance time with interactive systems. *Commun. ACM* 23, 7 (July 1980), 396-410.
3. Cockton, G. Getting there: six meta-principles and interaction design. In *Proc. CHI '09*. ACM press (2009), 2223-2232.
4. Cockton, G. Design Situations and Methodological Innovation in Interaction Design. *Ext. Abstracts CHI 2010*, ACM press (2010), 2745-2754.
5. Cockton, G. Making Designing Worth Worth Designing (2012). Available through: ii.tudelft.nl/ValuesInDesign/submissions/cockton.pdf
6. Desurvire H., Caplan M., Toth J.A. Using heuristics to evaluate the playability of games. *Ext. Abstracts CHI 04*, ACM press (2004), 1509-1512.
7. Furniss, D. Beyond Problem Identification: Valuing methods in a 'system of usability practice'. PhD Thesis. UCL. (2008). Available through http://dl.dropbox.com/u/2191887/My Publications/2008_Furniss_PhDThesis.pdf
8. Gaver W., Dunne T., Pacenti E. Design: Cultural probes. In *interactions* 6, 1 (January 1999), 21-29.
9. Law E. L.-C., Hvannberg E., Vermeeren A.P.O.S., Cockton G., Jokela T. Made for Sharing: HCI Stories of Transfer, Triumph & Tragedy. *Ext Abstract CHI'13*. ACMPress (2013).
10. Mankoff, J., Dey, A.K., Hsieh, G., Kientz, J., Ames, M., Lederer, S. (2003). Heuristic evaluation of ambient displays. In *Proc. CHI 03*, ACM (2003), 169-176.
11. Nielsen J. and Molich R. Heuristic evaluation of user interfaces. In *Proc. CHI '90*, ACM (1990), 249-256.
12. Rogers, E. M. *Diffusion of innovations.* New York: The Free Press (1983)
13. Woolrych A., Hornbæk K., Frøkjær E., Cockton, G. Ingredients rather than recipes: A proposal for research that does not treat usability evaluation methods as indivisible wholes. *IJHCI* 27(10), (2011) 940-970.
14. Cockton, G., A Load of Cobbler's Children: Beyond the Model Designing Processor. To appear in CHI EA 13.

Diffusion of Worth Mapping: The Worth of Resource Functions

Gilbert Cockton
Northumbria University
Newcastle upon Tyne, NE1 8ST, UK
Gilbert.Cockton@northumbria.ac.uk

ABSTRACT
This workshop paper uses a resource function vocabulary from the Working to Choose framework to analyse diffusion of the Worth Maps approach across several application domains. It explores how a resource function vocabulary can indicate aspects of design approaches and their use that favour successful diffusion.

Author Keywords
Diffusion of Design Methods; Worth Mapping; Working to Choose framework.

ACM Classification Keywords
D.2.2 Design Tools and Techniques

INTRODUCTION

Diffusion of innovation is a well understood phenomenon [9], but its insights have yet to be applied to specific interaction design and evaluation methods. This paper examines the roles of resource functions in the diffusion of Worth Mapping as an approach for interaction design, This paper applies selected insights from a companion paper for this CHI 2013 workshop, *Facilitating the Take-Up of New HCI Practices: a 'Diffusion of Innovations' Perspective*.

THE WORKING TO CHOOSE FRAMEWORK (W2C)

The Working to Choose (W2C) framework [4] is a conceptual system that: guides audit of existing approaches; identification of gaps in a design team's interaction design practices; adaptation and extension of existing approaches; and invention of new approaches. W2C analyses can focus on a single approach or an integrated group.

A design or evaluation *approach* is a group of *resources* [10] that seeds the *methods* that result from design work. Re-use of approaches requires adaptation, extension and completion to form viable methods. In this sense, methods are achievements, not premonitions. Design work can appear to be structured by methods, but the full details of methods in use cannot be known in advance, and certainly not before an approach is applied. Different approaches make different demands on design teams in terms of the work needed to make methods work, i.e., to become viable in specific work contexts.

W2C is a conceptual system for analysis of design and evaluation approaches, with three interlocking concept sets:

1. Resource functions
2. Design Choice Types
3. Meta-Principles for Designing

Recent developments have subsumed the last two concepts within the first, resulting in a simpler framework.

Resource functions

Resource *functions* are a reconceptualization of resource *types* [11]. In [11], resources within approaches were conceptualised as having distinct *types*, but it was later realised that a single resource can perform multiple types of *functions*, as well as having further attributes. The current resource functions (as named as types in [4]) are:

1. Directive
2. Harvesting
3. Expressive
4. Performative (additional to [4])
5. Scoping
6. Axiological
7. Knowledge
8. Invigorative (additional to [4])
9. Protective (additional to [4])
10. Integrative (additional to [4])

This mixes two vocabularies for resource functions, the original type vocabulary (1-3, 5-7) from [4] and an challenging vocabulary (4, 8-10) from [6], which presents further alternative vocabularies (everyday, technically neutral, poetic). Names from [4] have been used for consistency with the companion paper for this workshop. The provision of multiple vocabularies in [6] is intended to promote thought and reflection on resource functions, rather than create rapid but limited understandings through a single set of clear but incomplete definitions.

Resources can have multiple functions. For example, sketches have expressive, inquisitive, and directive functions. They not only express design ideas, but can also trigger inquiry and direct refinement. Also, sketches can have *performative* functions when shared with design

stakeholders. Sketching however is not guaranteed to have all four functions. It is always *expressive*, but must be used in particular ways to have other functions.

Three resource functions form the core of approaches: *integrative*, *scoping* and *axiological* (details below). As a shorthand, we can still write as if resource functions were types, i.e., when we call something a *scoping resource* this is to be understood as a resource with a scoping function.

Resource functions that were added since [4] recognise the complex, social, and emotional aspects of design work [6]. *Integrative* functions co-ordinate design approaches across design activities. *Performative* functions address communication beyond the design team through high quality visual presentations, careful use of language, and other presentation and communicative skills. *Invigorative* functions drive the design process forward, while *protective* ones keep it on track. Both are emotional functions.

Resource Functions Simplify and Widen Analysis

Design approaches have scopes that can be specific to technologies, user groups, application domains, or development roles or process stages. An approach's scope can also support one or more *Design Choice Types*, i.e., beneficiaries, purpose, artefacts and evaluations [4]. Making choices about beneficiaries can be a wholly human-centred activity, whereas making choices about artefacts can be creatively and/or technically focused. Making choices about evaluations can be human and/or technically focused, whereas making choices about design purpose can have a range of foci. W2C can thus be simplified by subsuming Design Choice Types into scoping functions.

W2C can be further simplified by treating Meta-Principles for Designing as evaluation criteria for resource functions [4]. For example, the quality of directive functions can be evaluated via the *tenacity* meta-principle. Directive functions are successful when they result in design options that remain valid choices due to the quality of design work that produced them. Options are not just potential features, capabilities and qualities of interactive digital artefacts, but also stakeholder considerations, design purposes and evaluation practices.

W2C's simplification subsumes two existing set of abstract concepts within ten currently identified types of functions that support a wider range of analysis. The main aim of this workshop position paper is to illustrate how a simpler W2C allows resource function analysis to be applied to an existing design approach, using *worth mapping* as an example. The use of worth mapping in six design contexts [1,2,3,7,8,9] is used to illustrate the balance between resources provided by worth maps and those that had to be sourced within specific project contexts. This exposes the role of local resources in the transfer of approaches from one design context to another, as well as demonstrating the role and value of the re-usable resources provided by design approaches. This widens analysis of the role of approaches in design work, as will now be illustrated via a resource function analysis of the worth mapping approach [3].

WORTH MAPPING

Worth Mapping as an approach creates Worth Maps, its main expressive resource, within a context of scoping and axiological functions (axiology is the study of values). We now briefly present worth maps and their use in design research settings since 2007, with published examples of use from the UK, Finland, Portugal, and Switzerland, in a wide range of sectors, including home systems, entertainment, ambient displays and mobile applications.

Worth Maps are box and arrow diagrams that represent relationships between artefacts, user experiences and design purpose. The term artefact is used here to refer to any designed product or service at any stage of realization from initial ideas to installed user bases, Boxes in worth maps either represent attributes of artefacts, episodes of interaction, or usage outcomes. Sequences of arrows between boxes can be followed from artefact attributes via interaction episodes to usage outcomes, creating *means-end chains* that represent intended, perceived or observed causal relationships between artefacts, user experiences and usage outcomes. Artefacts and experiences are means to ends, the latter correspond to worthwhile usage outcomes.

Reconstructed Applications of Early Worth Maps

First use of worth maps was based on reverse engineering from the design experiences of the author, who invented worth maps. Early versions of worth maps (W/AMs) were applied to e-commerce and educational examples [2], and to experiences from a student design exercise [3].

First Use on a Live Project

Figure 1 shows a worth map from early work within a Family Archive design research programme [7]. The inventor of worth maps was a permanent member of the research team for three months, during which time worth maps were restructured as a result of relating potential features of a family archiving system to valuable outcomes.

Consultancy Support for Two Live Projects

Worth Maps were used without direct involvement of their inventor on some case studies within the Finnish VALU project (including one on on-line gambling [8]), as well as an academic research project on ambient displays in Portugal [9]. Both projects used the revised worth map structures from the family archive project. The inventor provided advice via email, mostly on worth map element types, but this was far less significant than the local innovations within the VALU project [8]. For example, blank typed elements (e.g., for features or user experiences) were also used to good effect to indicate gaps in mean-ends chains for the current system.

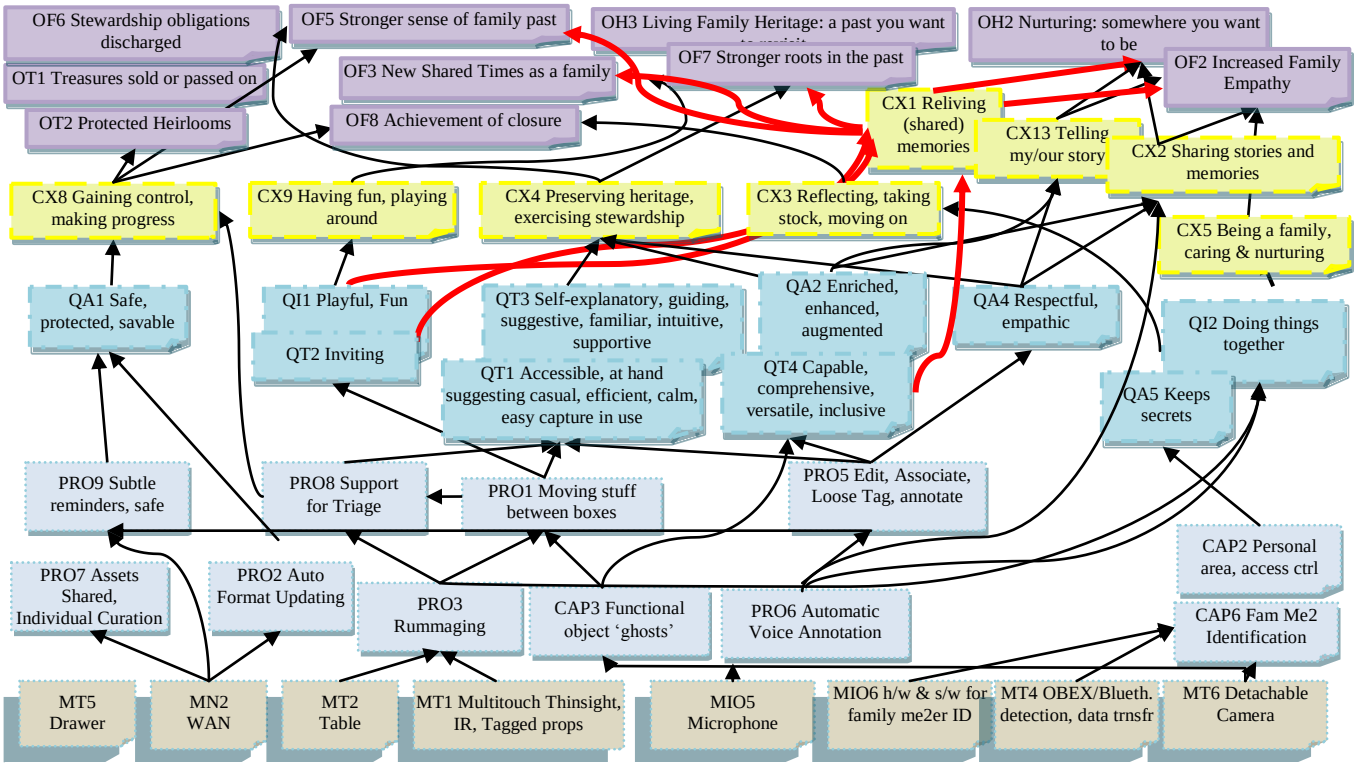


Figure 1: Example Worth Map (from [7])

Each box represents a specific type of worth map *element*. The purple boxes with O labels at the top represent *worthwhile outcomes* that were expected when organizing objects in a family archive. The green CX labelled boxes are *user experience* elements corresponding to episodes of interaction All elements below them are *artefact attributes*, separated here into *materials* (bottom M labelled layer), *features* (PRO and CAP labelled elements) and *qualities* (Q labels). Artefact features and qualities are linked to worthwhile outcomes via user experiences**,** since it is users who actually create worth through use. Bold arrows indicate such means-end chains.

Independent Use

Worth maps have most recently been used on a mobile application research project in Switzerland [1]. The inventor had no contact with this research until the completion of an associated PhD thesis [1]. Worth maps for complete design projects can become very large, and need to be modularised to manage complexity. In [1] worth maps were modularised around mobile phone elements, as this was a focus for the local mobile HCI research.

THE ROLES OF RESOURCE FUNCTIONS IN THE DIFFUSION OF WORTH MAPPING

All types of resource function have contributed to the diffusion of worth maps, but additional local project resources have been just as critical to success, e.g., the use of a diagram editor to layer worth maps in [8], and local identification of design and value elements [1,7,8,9]. This is consistent with a key position on approaches and resources, i.e., that virtually no approach has a complete set of resources prior to use. Approaches only become workable methods through local adaptation of their provided resources and local addition of additional resources. In [7], stages in the life cycle of an archived object were the basis for modularisation (a worth map for each stage), whereas in [8], worth maps were modularised around user experience, and in [1] around materials. Such variations show the role of local values and insights in adapting different structures for worth maps. The relation of local resources to those provided by worth mapping approaches is now reviewed.

Expressive Resource Functions

Worth maps are primarily *expressive* resources that compose a family of element types. These evolved rapidly during 2007, as shown in Table 1. The first set of element types for worth maps is shown in the left column. These revised the element types of previous worth/aversion maps (W/AMs [2]), which in turn were based on *hierarchical value models* (HVMs - a diagramming format with associated means-end chain and laddering theory) from marketing research. The main applications of HVMs included digital service research [2]. W/AMs revised HVM elements, in response to experiences of reverse engineering [2] and from a VALU project workshop before [8].

The key point here is that the evolution of worth maps' expressive function was not a process of linear progress, but one of contextual diversification where alternative worth map elements were adapted to different design contexts.

Table 1. Rapid Evolution of Worth Map Elements, 2007

Worth Maps	W/AMs	HVMs	Focus
Worthwhile Outcome		*Terminal Value*	Design Purpose and User Experience
User Experience	Usage Consequences	Instrumental Value	
Feeling		Psychosocial Consequence	
Action		Functional Consequence	
Quality	Abstract Product Attribute		Artefact
Feature	Concrete Product Attribute		
Material			

The first generation of worth map elements is shown in Table 1, left column. HVM concrete product attributes were divided into materials and features. W/AM compression of three steps in HVM means-end chains into a single usage consequence was replaced with more complex causal structure, only to be replaced in the second generation of worth map elements by a single user experience element type (see [7] for rationale). Figure 1 uses this set of element types (i.e., not one in Table 1). A third generation of worth map elements was proposed [8] that merged artefact elements into a single 'product attribute', resulting in only three types of element: artefact, experience and outcome.

Worth mapping has thus used five different, but overlapping, sets of element types. Worth Sketches (boxes but no arrows [3]) are a further alternative expressive resource. Worth mapping has thus shown much expressive variation. As well as differences in structure and content, worth maps have been modularized [1,7] and layered [8] differently. Although it was not known at the time, worth maps were developed through a series of collaborative case studies as *approaches* [11], where resources were adapted to project circumstances and experiences of their use. Thus each set of elements has demonstrated different (but overlapping) benefits and challenges in specific settings.

First generation worth maps only lasted months, but the second generation (feelings and actions become parts of user experiences) achieved some successes with multi-disciplinary R&D teams including hardware and software engineers, interaction designers and human science specialists [1,7,9]. However, this simplification was not enough for a product development team that only included one technical role [8], which motivated the proposal to have only one type of artefact (product) attribute. In all cases, no-one concluded that worth mapping could not transfer to their design setting. Instead, element types were successfully adapted to better fit project contexts. The diffusion of worth maps thus depended on local adaptations and extensions that exploited the possibility of any use of physical cards or drawing tools (with/out layering) for any set of element types being valid as long as it is compatible with worth mapping's values (*axiological* functions, next).

Scoping and Axiological Resource Functions

Scoping resources limit an approach to specific abstract design situations or development process stages. The scope of worth mapping is limited to the range of choice types that design teams choose to co-ordinate via them (see *integrative* function below). Worth maps should thus transfer to any design context. Indeed, they have been found to be valuable for marketing as well as design [1,8]. No restrictions have yet emerged in terms of application domains addressed to date [1,2,3,7,8,9]) or in terms of the technologies involved (domestic controls [3], ubiquitous computing [7,9], mobile phones [1] and web-based services [2,8]). Worth maps have been used across complete development lifecycles [1] and roles involved in their use have included marketing, finance, software and hardware engineers, interaction designers and human scientists. A current case study is focused on a social network for the care circles of children with major impairments.

Despite this wide coverage across over six usage contexts, *axiological* functions inevitably limit worth mapping to project contexts where explicit links between design features and stakeholder benefits are valued. Design settings where such links can remain tacit are thus not in scope for worth mapping approaches, and thus will not transfer there. Similarly, worth maps are intended for acentric design processes, i.e., ones that have no single centre, but instead shift foci and emphases between different types of choices and their co-ordination [5]. By not privileging one type of choice, as user-centred design privileges user *beneficiaries*, *acentric* design processes must *balance* and *integrate* different design foci [5]. During design, the focus can shift between all four choice types, i.e., the artefact, beneficiaries, evaluations and purposes.

Worth maps also value expression of design purpose as intended worth in the world, but are neutral on whether this should be wholly grounded in empirical data on users' wants and needs (although it can be). This lets design teams offer unexpected value and experiences that beneficiaries do not currently know are possible. Design teams are thus allowed to be *generous*. As well as specific commitments to explicit purpose as worth, and links between this and artefact attributes, *balance*, *integration* and *generosity* are core values for worth mapping [5], which also values design process freedom where options for worth map elements can be developed in any order. Overall, the openness and freedom of worth maps has eased their diffusions as design practice innovations.

Directive, Harvesting and Integrative Functions

Directive functions guide use of an approach. Only one simple worth map construction procedure has been published by the inventor [3]. A local directive function

evolved in [7] as a combination of collaborative card based worth sketches that the inventor subsequently turned into a digital worth map. In [1,8,9], design teams used their own knowledge, experience and insights to direct worth map development. A core objective for [1] was to develop structured approaches to worth map creation, so there are now two alternative published directive resources [1,3] for worth mapping, as well as partially documented local directive resources [7,8,9]. Local creation of directive functions have thus been possible, as has creations of re-usable ones as resources for worth mapping approaches.

Worth Mapping has no *harvesting* functions to provide sources of data and inspiration, but project teams can find complementary approaches to compensate, including field research and design workshops [7], online sentence completion and existing product attributes [8], and interviews and competitor analysis [1]. As long as harvesting resources deliver design purpose elements at appropriate levels of generality, worth mapping can use any user research or design ideation approaches.

Worth maps primarily have an *integrative* function, and can co-ordinate activities focused on design purpose and artefact design, as well as activities resulting in understandings of beneficiaries (via user experience elements in worth maps) and also evaluation activities (by associating measures and targets with worth map elements). A range of scoping functions results in worth maps co-ordinating two, three or four types of design choice. Element types also have scoping functions, with the simplification to three element types in [9] motivated by a predominance of non-technical roles in the project team. While technical and creative specialists can make good use of the full range of five element types, non-design roles can find them overly complicated. Worth map element types thus have a scoping function that can adapt to the capabilities and preferences of design team members.

Performative Resource Functions

Performative resources support communication and persuasion within design settings. Physical worth sketching cards [7] can be regarded as a *performative* resource that shares a current set of worth map elements within a design research team. Layering worth maps using a drawing editor [8], limiting visible layers to those involving one or more specific user experience elements, proved to have a useful performative function when communicating worth maps to audiences who were not involved in their creation. Even so, it proved difficult to communicate completed worth maps to those who had not contributed to making them. Still, the project team in [8] was multidisciplinary, involving sales, marketing and customer relationship management, and worth mapping did improve communication between these roles. Initial difficulties did not prevent the project team from presenting insights from worth mapping to several national divisions. Additional local performative resources here were key to presenting focused insights. Such local resources may prove vital to improving diffusion in contexts where there is diverse disciplinary expertise.

Emotional Resource Functions

Emotional resources have rarely been given attention in design method research [6]. Emotional resources can have invigorative or protective functions. Resources with an *invigorative* function accelerate the progress of design towards successful completion. In [7], group use of worth sketching cards created tactile social experiences that enlivened design discussions and supported team creativity and critique that respected and exploited the expertise of each team member, creating common ground across the team (this was also achieved in [1,9]). A focus on value innovation through worthwhile outcomes accelerates progress in design workshops and generated valuable new design opportunities [1,7,9]. Users' motivations were effectively represented in a compact format (outcome elements). The final worth mapping document in [7] provided value within for at least 18 months after the initial worth mapping, and the value/outcome elements identified have guided subsequent design research. In [8], the value elements made it easier to prioritise the existing backlog in an agile development context. In [8], worth maps also provided a valuable additional high level viewpoint on evaluation of business value, but this was in part due to the involvement of product and financial managers for a live commercial service. These roles provided local invigorative resources that were vital for success.

Invigorative resources give design teams confidence that their design work is worthwhile, allowing them to focus on adding further value. In contrast, *protective* resources keep design on track, avoiding dead ends and unproductive debates, and thus reducing the costs of adverse outcomes during design. The protective function of worth maps improved as element types stabilised. Difficulties of understanding in [7,8] did not arise in [1]. Protective resources depended on the inventor in [7] and careful management of relationships beyond the project team in [8]. There was no evidence of protective resources for [1,9].

Emotional resource functions are emergent, and always due to interactions between an approach's resources and their specific configuration and use in design settings.

Knowledge Resource Functions

Knowledge resources provide underpinning knowledge, concepts and theory that enables design teams to make best use of approaches. A basic grasp of laddering, means-end theory and consumer psychology, marketing and related applications of. HVMs were achieved via different local resources across the projects. In [7], a key local resource was an evolving tutorial document written and updated for the project team by the inventor. In [8,9], it was coaching and advice by email. In [1], it was the lead researcher's interest in worth maps that motivated her to review the relevant literature on laddering, means-end theory, HVMs

and related consumer psychology and marketing concepts. The inventor's ability to respond in [7] to problems of worth map complexity depended on his familiarity with known issues with HVMs in the marketing literature [2]. Knowledge resources are vital to successful adoption and adaptation of approaches. Full competence in the use of directive and expressive resources depends on them.

DIFFUSION OF INNOVATION CONSIDERATIONS

The Diffusion of Innovation literature [10] provides many insights into why specific functions from worth mapping's resources and also local project resources combined to result in successful design work. The companion paper presents these insights systematically. A few are now presented for worth map diffusion.

The *trialability* (fourth attribute of innovation [10]) of worth maps was aided by research context values (local axiological resources) that allowed experiment, provided missing harvesting and directive resources, and even extended scoping functions to include marketing uses [1,8]. For the latter, local marketing knowledge resources [1,8] and familiarity with HVMs and consumer psychology [1] reconnected worth maps with their origins [2].

Homophily (change agent success factor: shared beliefs and team attributes [10]) let appropriate harvesting and directive resources be provided locally, with quantitative approaches used in [8], qualitative in [7], creative brainstorming in [9] and engineering requirements approaches in [1].

The minimal scoping function of worth mapping (a choice of co-ordinating 2, 3 or 4 choice types) means that only axiological mismatches at the *organisational decision making stage* [10] can put worth mapping out of scope. Local and secondary *knowledge* resources extended worth mapping's scope to marketing and communications [1,8].

As regards the *persuasion/interest stage* of innovation [10], axiological resources attracted design teams here. In [1], a strong local need advanced worth mapping to the *routinizing* stage [10] via new directive resources. In contrast, in [7,8], new local axiological resources removed the need for the expressive and integrative functions of worth maps. What diffused here instead was the worth-focused context [3] within which worth maps had been developed. After use on one project, worth maps were no longer needed to maintain a value focus, but nevertheless, the expressive functions of worth maps was a factor in the initial *agenda setting stage* [10] that persuaded design teams of a match to their needs or aspirations.

CONCLUSIONS

Overall, local resources were critical to the success of worth mapping, as predicted by [11]. Concepts from W2C may thus be combined with diffusion of innovating findings to explain the success of worth mapping across a range of project contexts. Core *expressive*, *integrative* and *axiological* functions of worth mapping resources were unchanged in [1,9], but were reduced/simplified in [7,8]. All projects [1,7,8,9] successfully added appropriate (i.e., *homophilous* [10]) *directive* and *harvesting* resources that are intentionally missing from worth mapping. Although performative functions were unintentionally omitted, all projects managed to add these locally.

Future work with worth mapping needs to pay more attention to resource functions identified since [4] (performative, invigorative, protective), perhaps adding new resources to provide more re-usable support for the social and emotional aspects of design work [6]. This would further demonstrate the worth of resource function analysis.

REFERENCES

1. Camara, F.G. 2012. *Retours d'expérience sur la conception centrée valeur de Cocoon vers des arbres de vie*. PhD Thesis, University of Grenoble.
2. Cockton, G. 2008. Putting Value into E-valu-ation. In *Maturing Usability: Quality in Software, Interaction and Value*, eds. E. Law, E. Hvannberg and G. Cockton, 287-317, Springer.
3. Cockton, G. 2008. Designing Worth: Connecting Preferred Means with Probable Ends. *interactions* 15(4), 54.
4. Cockton, G. 2012. Making Designing Worth Worth Designing. Available through: ii.tudelft.nl/ValuesInDesign/submissions/cockton.pdf
5. Cockton, G. 2012. UCD: Critique via Parody and a Sequel. *CHI 2012 Extended Abstracts*, 1-10.
6. Cockton, G., A Load of Cobbler's Children: Beyond the Model Designing Processor to appear in *CHI 13*.
7. Cockton, G., Kirk, D.S., Sellen, A. and Banks, R. 2009. Evolving and augmenting worth mapping for family archives. *Proc. BCS HCI 2009*, 329-338.
8. Cockton, G. Kujala, S., Nurkka, P. and Hölttä, T. 2009. Supporting Worth Mapping with Sentence Completion in *Proc. INTERACT 2009*, Part II, 566-81.
9. Otero, N., and José, R. 2009. Worth and Human Values at the Centre of Designing Situated Digital Public Displays. *Int. J. Advanced Pervasive and Ubiquitous Computing*, 1(4), 1-13.
10. Rogers, E.M. 1983. *Diffusion of innovations.* Free Press.
11. Woolrych A., Hornbæk K., Frøkjær E. and Cockton, G. 2011. Ingredients rather than recipes: A proposal for research that does not treat usability evaluation methods as indivisible wholes. *IJHCI*, 27(10), 940-970.

Comparison of Case Studies of Usability Maturity Assessment and Process Improvement

Nigel Bevan
Professional Usability services
London, UK
nigel@nigelbevan.com

ABSTRACT
This paper describes a new case study of the implementation of usability maturity assessment and usability process improvement in three sites of a multi-national company, and compares the results with previous case studies.

In the new case study, a usability manager followed published good practice to successfully create a business case for human centered design, assess usability maturity, and select methods to implement process improvements.

GOOD PRACTICE IN MATURITY ASSESSMENT
The purpose of a usability maturity assessment is to profile the capability of an organization to take account of human-centered issues in all relevant design, development and support activities. By identifying the strengths and weaknesses of an organization it is possible to identify potential areas for improvement, and to suggest the most cost-effective methods and techniques that could be used to improve the capability.

ISO TR 18529 "Human-centred lifecycle process descriptions" [9] provides a model of good practice that can be used for assessment, and is divided into seven processes:

1. Ensure HCD content in system strategy
2. Plan and manage the HCD process
3. Specify the stakeholder and organizational requirements
4. Understand and specify the context of use
5. Produce design solutions
6. Evaluate designs against requirements
7. Introduce and operate the system

Each process is composed of a set of activities needed to achieve the objectives of the processes. The 43 activities in the model act as a checklist of good practice in human centered design. The model is tailored to the needs of an organization before use, eliminating any processes and/or activities that are either not relevant to the business, or outside the scope of the assessment.

Assessment is carried out by an auditor who is experienced in process assessment, the ISO 18529 model, and user centered design. The auditor also needs sufficient understanding of the business objectives and the design and development processes of the organization being assessed, to understand the potential business value to the organization of each activity.

The results of the usability maturity assessment can be used to identify cost-effective methods and techniques that can be integrated with the existing system development processes to improve the usability capability.

MULTI-NATIONAL COMPANY CASE STUDY
The usability manager of a multi-national European company believed that the poor usability of the company's public facing web sites were restricting business volume and profitability. He obtained a budget to benchmark the usability of the sites in several of the countries where the company operated, and to compare them with competitor sites. Satisfaction was measured by surveying a sample of the company's customers, and usability testing was used to identify problems and obtain measures of performance.

In the survey, many customers reported that they were unable to achieve all their objectives when using the web sites, and usability testing confirmed poor performance on some of the key web site tasks.

The usability manager estimated the economic benefits that could be obtained from improved usability, and on this basis obtained a further budget to assess the usability maturity and to facilitate usability process improvement at the offices in the surveyed countries.

USABILITY MATURITY ASSESSMENT
The assessments were planned and moderated by a team composed of the corporate usability manager, a representative of a usability agency that had worked with the company, and myself as a usability process improvement expert.

Processes 2-6 in ISO 18529 were relevant to the assessment. The model was also adapted to include some of the more recent items from the more recent and comprehensive ISO 18152 [7], resulting in a total of 33 activities.

The first step in each country was a preliminary visit to hold a 2-hour workshop to explain the corporate strategy for usability process improvement, the role of the usability maturity assessment workshops at the site, and who should participate.

A usability champion was identified at each site, who was given responsibility for promoting usability in their organization. There had been concern that the usability maturity assessments could be seen as an imposition by head office: sending in a team to identify what was wrong with the development process at each site. However, the reaction in each organization that we visited was reassuringly positive: the development teams expressed a desire to produce more usable web sites, even though in most cases, they had no idea how to do this. They thus appeared to perceive the visit as educational, empowering them to carry out activities that previously did not seem feasible. Basing the improvements on internationally agreed standards for good practice also helped provide credibility.

PROCESS IMPROVEMENT WORKSHOPS

Each assessment consisted of 5 half-day workshops, one for each process being assessed. Relevant stakeholders from management, design, development, marketing and support were invited to each workshop as appropriate, with 3-5 stakeholders typically participating in each. Almost all those who were available on the day accepted the invitation to participate, although at some sites it was difficulty to engage anyone from marketing.

Level 0	Incomplete (not able to carry out process)
Level 1	Performed (individuals carry out process)
Level 2	Managed (quality, time and resource requirements for process known and controlled)
Level 3	Established (process carried out as specified by organization, resources are defined)
Level 4	Predictable (performance of process within predicted resource and quality limits)
Level 5	Optimizing (organization can reliably tailor process to particular requirements)

Table 1. Maturity levels in ISO 15504-5 [7].

Each activity was assessed, and agreement reached as to whether it was Not performed (< 10%), or Partly (< 50%), Largely (> 50%) or Fully (> 90%) performed, and if Largely or Fully performed, the level of maturity on a scale of Performed, Managed, Established, Predictable to Optimizing (Table 1) [6]. A target to be achieved was also agreed, with Established identified as a desirable long-term goal.

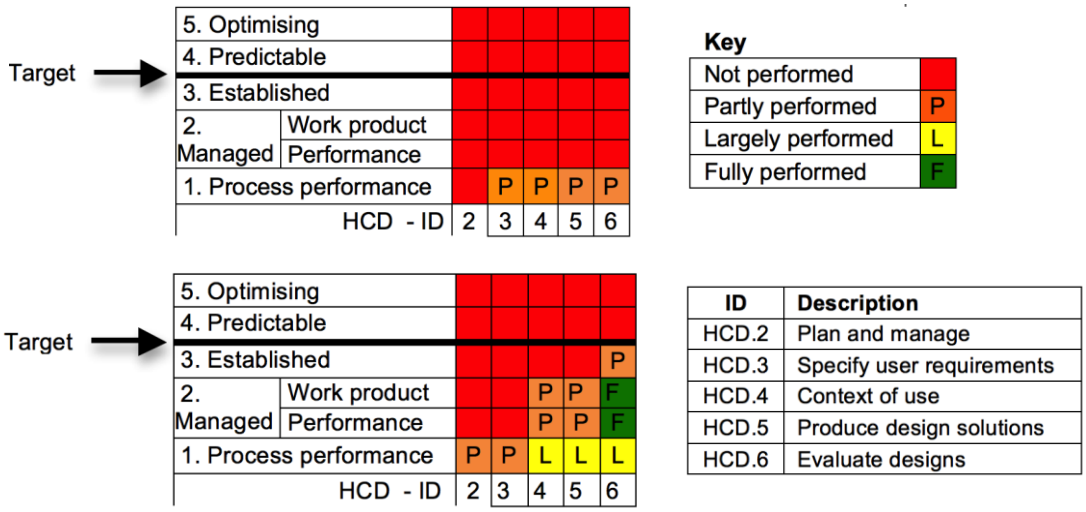
A preliminary summary of the results was reported back to local management in a final session.

REPORTING THE RESULTS

The full report, presented on a later visit, included a graphical overview of the results. Examples of the different types of results that can be obtained from usability maturity assessments are shown in Figure 1. The first set are the type of results that would be expected for an immature organization which was carrying out some ad hoc technical activities that contributed to user centered design. The second set of results are typical of an organization that is experienced at usability evaluation but less consistent in identifying the context of use and incorporating usability in the design process, and weak on specifying requirements.

The report also contained a more detailed breakdown of the results in each area (such as that in Table 2), together with an explanation of the factors that contributed to the rating and detailed recommendations for improvements.

The report concluded with a summary of the process improvement recommendations, together with the potential business benefits and suggested actions. Of 61 recommendations across 3 sites, 61% suggested using specific types of methods (such as personas and scenarios of use), 23% proposed activities (such as usability bug tracking or reviewing the risks resulting from poor usability), while 16% related to more general processes (such as assessing when user involvement would be beneficial or using a wider range of UCD methods).



ID	Description
HCD.2	Plan and manage
HCD.3	Specify user requirements
HCD.4	Context of use
HCD.5	Produce design solutions
HCD.6	Evaluate designs

Figure 1. Typical examples of usability maturity results.

ID	Name	Now	Plan
3.1	Describe the objectives which the user wants to achieve through use of the system	P	L
3.2	Understand the differing requirements of the different user groups (and other stakeholders) affected by the system	P	L
3.3	Research and agree required system usability, including expected behaviour and performance of the system with respect to the users	P	L
3.4	Define measurable criteria for assessment of usability in the intended context of use	P	L
3.5	Identify the business implications of the user requirements	L	L
3.6	Analyse the user requirements and develop an explicit statement of the user requirements for the system	P	L
3.7	Present these requirements to project stakeholders for use in the development and operation of the system	L	L
3.8	Evaluate the extent to which usability criteria and other human-centred requirements are likely to be met by the system and take effective mitigation to address risks to system usability	L	L
HCD.3	**Specify the user and organisational requirements**	**P**	**L**

Table 2. Assessment of activities.

Organizational characteristic	IAI	Inland Revenue (IR)/EDS	Multi-National sites
Wake-up call	Complaints of poor usability by customers	Lessons from usability testing	Poor usability found from surveys and testing
Executive champion	Quality manager	Senior management	Board support
Methodology	Based on ISO 13407	Based on ISO 13407	Based on ISO 9241-210
Training	On the job	On the job	On the job (planned)
Showcase projects	Yes	Yes	Planned
Attitude to process improvement	Committed	Committed	Sympathetic
Use a fully documented process?	No	Yes	Yes
Importance of end user needs	High	High	Moderate
Experience with usability	Moderate	None	None-Moderate
Number of stakeholders interviewed	8	13	5-16
Number of activities judged relevant and assessed	33	39	33
Number of days of interviews/workshops	1.0	6.0	2.5
Proportion of activities largely or fully performed	27%	51%	24-52%
Number of new methods recommended to be used	10	8	7-16

Table 3. Comparison of assessments.

TRANSFERABILITY OF ASSESSMENT AND IMPROVEMENT METHODS

This section compares previous case studies of usability maturity assessment and usability process improvement at IAI and IR/EDS [2,1] with those at the multi-national sites.

Take up of usability process improvement

Schaffer [15] states that the prerequisites for institutionalization of usability are: a wake-up call to motivate change, an executive champion to ensure organizational commitment, a methodology to implement User Centered Design (UCD[1]), training in UCD methods, and showcase projects to demonstrate the benefits. The role

[1] UCD is used in this paper to refer to both User Centered Design, and the similar ISO concept of Human Centered Design.

of these factors in the 3 case studies is summarized in Table 3.

All the organizations were concerned about the usability of their products or services. The multi-national had Board support for the assessments, but the improvements at IAI and IR/EDS were carried out as part of an EU research project, and had less senior involvement. This may explain why in both of the latter organizations, the first project nominated as the showcase to trial the UCD methods dropped out. In both cases the project manager decided that the claimed benefits were outweighed by the potential risks of using new methods. But both organizations found replacement projects, and the eventual benefits were judged to be so cost-effective that both organizations decided to incorporate the methods as part of their documented systems development methodology.

Additional factors that Bevan [2] identified as contributing to the success of the IAI and IR/EDS process improvement activities was the culture of both organizations to:

- provide systems that meet user needs
- improve their processes.

While neither of these commitments was as strong in the multi-national sites, all the sites welcomed the opportunity to improve their processes with the support of head office.

Other approaches to maturity assessment

There are not many other published case studies of usability process improvement. Jokela [14] has reported less success. It is possible that there was less management support for Jokela's assessments. Another potential difference could be in the way that the assessments were carried out. In the multi-national assessments, the initial discussion of the extent to which each specific activity had been carried out was quite short, but stimulated a much longer discussion of the more general problems associated how the activity could be achieved within the context of the organization. Thus the assessments of each activity were based on a broader understanding of the context in which it took place. As Jokela notes [11] it is important that one verifies that the results of an activity are correct and sufficiently comprehensive, so that they have a genuine impact on the system under development.

Jokela's KESSU model [12] is for the technical content of the activities (level 1 in Table 1), rather than the extent to which they are managed (levels 2-5). Jokela says "I really believe that generally there are major problems and a lot to improve in the how companies do and understand the substance of usability. And examining levels 2 - 5 in an assessment is not very meaningful if there are problems in the substance." [11]. It is true that the ISO maturity levels were developed for the much more mature field of systems engineering, but they do become relevant once an organization is Largely performing a process (as in the second example in Figure 1).

Time needed for a usability maturity assessment

The characteristics of the multi-national sites (Table 3) covered a similar range to the previous assessments in terms of the existing experience of usability, the number of stakeholders interviewed, and the proportion of activities that were assessed to be largely or fully performed.

IAI was a smaller organization with a more informal culture which made it possible to carry out an initial lightweight assessment in 1 day, prior to subsequent activities to plan process improvement. The Inland Revenue/EDS requested a more rigorous formal assessment (similar in approach to a CMMI assessment [3]) with 2 assessors working in parallel for 3 days. At the multi-national sites, the 5 half-day workshops provided sufficient information to make process improvement recommendations.

Usability process improvement

In all 3 case studies, ISO 13407 (now ISO 9241-210) supported by the detail in ISO TR 18529 and ISO TR 18152 were used as the definition of good practice. All these standards define user-centered design objectives and activities, without stipulating specific methods.

In the case of IAI which had not previously used any usability methods, a complete methodology was proposed mapped onto the ISO processes (Figure 2). The methods used to implement the methodology were selected to be suitable for use by existing staff who had limited usability skills [1].

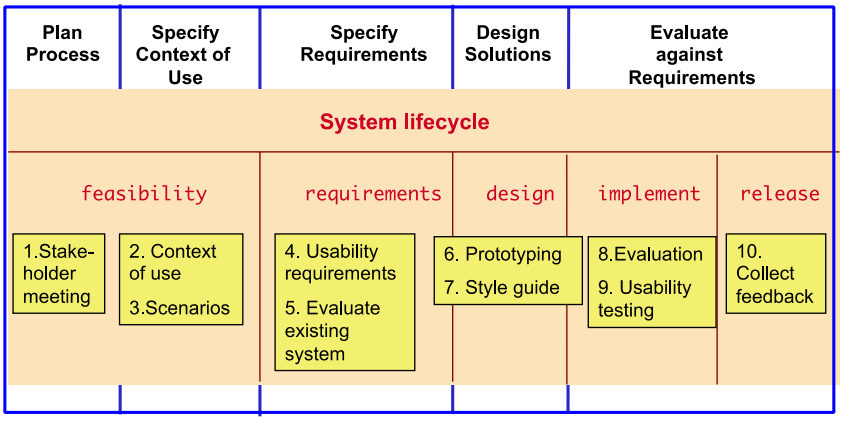


Figure 2. Methods recommended for use by IAI.

IR/EDS was already skilled in usability evaluation, but the assessment revealed that it needed to gather more information about the context of use early in development; produce scenarios of usage; establish usability requirements; use checklists and guidelines; evaluate early prototypes; and to improve the existing Joint Application Design (JAD) workshops to focus on task scenarios rather

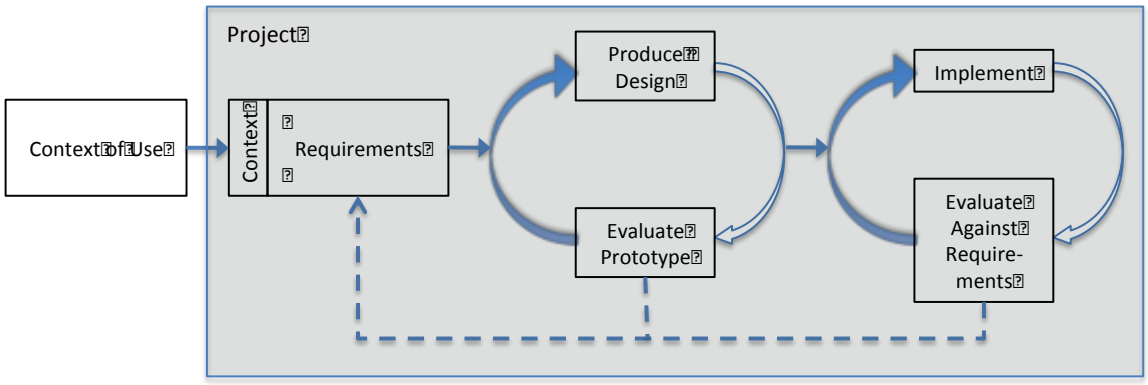


Figure 3. User centered design lifecycle.

than screens or functions, to give the participating users a stronger voice, and to use usability acceptance testing to validate the JAD design.

For the multi-national assessment, the relationship between the ISO processes and the system lifecycle was represented in a way that gave more emphasis to iteration, and did not identify any specific methods (Figure 3).

At the least mature multi-national site, the methods recommended for process improvement were: identifying the context of use from focus groups and market research data; deriving user requirements from context of use data, business requirements and other data; and iterative design creation using simple usability evaluation methods.

The next most mature site already carried out some heuristic evaluation and user testing. The same methods were recommended, together with site-specific needs for: use of surveys; designs that support multiple user groups and customization; use of a style guide; and use of usability walkthroughs.

The most mature site already used focus groups, personas, scenarios and usability testing. Additional methods recommended were: developing a checklist of potential stakeholders; developing a standard form for stakeholders to communicate suggestions; comparative/competitor usability testing; diary studies (or similar); deriving user requirements from context of use data; business requirements and other data; producing alternative design solutions, iterative design creation; accessibility testing or audit; and discussing the implications of the usability test results with the evaluators.

Local usability agencies were identified to provide training and support in use of the new methods.

CONCLUSIONS

The case studies provide examples of transfer that have been achieved by taking established good practice (supported by ISO standards) for:

- Usability maturity assessment
- Usability process improvement
- Choice of specific UCD methods

Implementation has been tailored for the different environments summarized in Table 3. All the organizations welcomed the opportunity for usability process improvement based on the results of the usability maturity assessment.

The processes in the ISO 18529 model structure the UCD activities into groups that appropriately mapped onto the responsibilities of particular subsets of stakeholders in the assessed organizations. However, the formalism of the model can make it difficult for the organization being assessed to understand the intended meaning. It is therefore important for the assessors to have enough knowledge of the organization to enable them to explain and interpret the model in the context of the organization's processes.

In order to ensure that the content of the model comprehensively covered all relevant UCD activities at the multi-national sites, additional activities were included from ISO 18152. The new standard currently being developed to replace ISO 18529 [10] is expected to include these extra activities.

From the experience of these assessments, the key success factors for implementation of the approach appear to be:

- Awareness in the organization of existing problems with usability.
- Senior management support for change.
- A desire by staff to improve.
- Basing the assessment on an international standard model for good practice.
- Assessors who can relate the model to the organizations' processes.
- A broad knowledge in the assessment team of potential methods and techniques that can be used to improve usability.

The IAI and IR/EDS assessments were carried out in 1999, and the multi-national case study in 2011. It is surprising that the range of usability maturity found in some organizations today has changed so little from what was found 12 years ago.

ACKNOWLEDGEMENTS

The high professional skills of my colleagues in the multi-national assessment team were a major factor in ensuring that the case study was implemented successfully.

REFERENCES

1. Bevan, N. (2005). Cost benefits framework and case studies. In R.G. Bias, & D.J. Mayhew (eds), *Cost-Justifying Usability: An Update for the Internet Age*. Morgan Kaufmann.
2. Bevan N, Earthy J (2001) Usability process improvement and capability assessment. *In Proceedings of Joint AFIHM-BCS Conference on Human-Computer Interaction IHM-HCI'2001, Volume 2. Vanderdonckt, J., Blandford, A. & Derycke, A. (eds.)*, Toulouse: Cépaduès-Editions.
3. CMMI Product Team (2006). *CMMI for Development, version 1.2*. http://repository.cmu.edu/sei/387/
4. ISO 9241-210 (2010) *Human-centred design for interactive systems*
5. ISO 13407 (1999). *Human-centred design processes for interactive systems.*
6. ISO/IEC 15504-3 (2004) *Information technology -- Process assessment -- Part 3: Guidance on performing an assessment*
7. ISO/IEC 15504-5 (2012) *Information technology -- Process assessment -- Part 5: An exemplar software life cycle process assessment model*
8. ISO TS 18152 (2010). *A specification for the process assessment of human-system issues*.
9. ISO TR 18529 (2000) *Human-centred lifecycle process descriptions*
10. ISO NP 9241-220 (2013) *Processes for enabling, executing and assessing human centred design within organizations*
11. Jokela, T. (2013) Personal communication.
12. Jokela, T. (2007). Characterizations, Requirements and Activities of User-Centered Design - the KESSU 2.2 Model. Maturing Usability. *Quality in Software, Interaction and Value.* E. Law, E. Hvannberg and C. Cockton. London, Springer-Verlag: 168-196.
13. Jokela, T., Siponen, M., Earthy, J. and Hirosiwa, N. (2006). A Survey of Usability Capability Maturity Models: Implications for Practice and Research. *Behaviour & Information Technology* 25(3): 263-282.
14. Jokela, T. & Iivari, N. (2001) Usability capability assessments - experimenting and developing usability maturity models. *Proceedings of IHM-HCI 2001*, Volume 2.
15. Schaffer (2004) *Institutionalization of Usability: A Step-by-Step Guide*. Addison-Wesley.

Interviews Rather than Field Studies: Better Understanding on What-Is-to-Be Developed with Less Efforts?

Timo Jokela
Aalto University & Joticon
timo.jokela@joticon.fi

ABSTRACT

Field study methods such as contextual inquiry are effective methods for understanding users' world. They, however, require typically a lot of resources and are expensive. In some of my projects I have not had such resources available. In these cases face-to-face interviews have been the choice as a less resource consuming and cheaper method. However, I think that the result has been almost 'a triumph'. I think that my 'substance interview' strategy helps to gain a very good understanding of what is to be developed – probably even deeper than field studies would provide. Further, the usefulness of the interview method does not seem to be context dependent, i.e. is 'transferable'.

INTRODUCTION

Usability is about supporting users' world. Systems that are used at work should to support users' work. Therefore, it is essential to understand users world and work as the basis of system development.

When I have done usability evaluations (expert evaluations), I have found almost regularly that some usability problems origin from the fact that the designers have not understood users' work. One could even state that the designers had not truly even understood what exactly is the object of the design. And these problems typically are both the most severe ones and difficult to fix.

Understanding users' work is a challenging task. Field study methods such as contextual inquiry (Holzblatt, 1993) are effective methods for gaining the understanding. They, however, typically require a lot of resources and thereby are expensive.

In many of my consulting cases, I have not had resources for field studies. Therefore, face-to-face interviews have been the choice as a less resource consuming and cheaper method. I have, however, found that one can gain a very good and deep understanding of users' world with interviews. Even to the extent that field studies would probably not provide much added value.

Probably this is not 'method transfer' case because interview is a different method than contextual inquiry. So, the message of this case study is that one always cannot 'transfer' a method one would like to use, due to the business context (money, resources). Therefore, does not allow using 'best' methods but use a 'cheaper' one.

But the other message is that I believe that I was able to develop an 'old' method (interview in this case) to a more effective one, just because of the reason that 'I had to'. And further, the experience so far indicates that this method is very transferrable.

I call this interview approach as *substance interview*. And my perception is that substance interviews work so well that I do partly question the necessity of field studies.

I think that this paper is a story of "triumph". However, this is conclusion is based on my personal impression and some informal feedback from the customers; not on disciplined research.

MAIN FEATURES OF SUBSTANCE INTERVIEWS

The goal is to understand 'what system is to be developed' essentially but deeply. So, actually this is a bit more fundamental analysis than understanding users' work.

The main features of substance interviews are:

Outcome

I have not exactly formulated how exactly call the outcome of the interviews. Customer feedback has been excellent. But when I ask customers what exactly is the thing that we produced, they cannot say.

What I can say is that output is an essential and deep model of the substance of the system to be developed. And the output is totally implementation independent.

Anyway, I model the outcome as a (large) mind map. The other, and a very essential outcome is the project team's increased understanding of the system, and the common shared terminology and language.

To Whom

I have used the approach in IT development projects. It, however, should be applicable more widely in other kinds of projects because the output is implementation free.

Benefits

The project team will know from the beginning what essentially and exactly we are planning to develop. Tacit knowledge will be made explicit, and undesired surprises will be avoided in the latter phases of development. Provides a common language to the project team.

When to do

The interviews are most useful to do in the very beginning of the project, before requirements definition. If done afterwards, helps to identify main usability problems. (This is the origin of the approach).

How to do

The interviews are carried out in half a day interview sessions. These are repeated until the system-to-be-developed is analyzed thoroughly enough. In between the interview sessions, I refine the model by myself alone.

Interviewees are those who have good knowledge of the application domain.

The driving force of the interviews is that I personally need to understand 'what is to be developed'. The philosophy is that if I understand the system-to-be-developed, then the interviewees and the project teem understand, too.

During the interview, I model on-line the 'what is the system' as mind maps on a shared screen.

Required resources

The resource needs depend, naturally, on the complexity of the system. The other factor is how easy it is to reveal the tacit knowledge: how well the interviewees can articulate what they know.

Based on the experience so far, the required resources are from three consulting days and up.

Application contexts

I have done these kinds of interviews on a number of different applications: a check-in system for a hospital; a government license management system, a commercial web application, an ERP system, and a healthcare directory system.

Customer feedback

The informal customer feedback has been most encouraging:

- "Viewpoints from all stakeholders incorporated"
- "The basis for specifications; we always come back to check things from the mind map"
- "One can keep track about the whole as well as details"
- "You can see visually the whole all the time"
- "Everything essential will be included"
- "A very useful way to make all parties to understand the things and terms in the same way"
- "An absolutely useful way to start a project and to make sure that all understands the things the same way.

CONCLUSION AND DISCUSSION

I briefly described the 'substance interview' method for understanding what is to be developed.

The origin of the method was due to method transfer problem: I did not have resources to do field studies and thereby not to transfer a method that I originally planned to use.

My solution was to do 'just interviews'. They, however, proved to be very effective ones. I believe that I developed a new kind of interview strategy that I call 'substance interviews'.

Now what is interesting from method transfer point of view, my experience indicates is that the usefulness of substance interview method seems not to be application depended at all. I have used it in projects of different application domains, and every time has worked.

I believe that the method is transferrable even outside ITC systems. The outcome of the interviews is descriptions of what to be built are totally implementation independent. I believe, but have no cases, that the method could be applied in every project which aim to develop 'something'.

Limitations

I have not really studied what other kinds of interview strategies exist. There may exist some that are close to mine that I am not aware of.

The usefulness of the method is based my informal judgments and limited customer feedback; not on objective research.

Implications for practice

Based on my experience so far, my absolute recommendation is to do this kind of interviews in every project (!), in the very beginning of the project.

The challenge naturally is that who would to this job.

New research items

There is absolutely need for research – actually, no true research as yet been done. In addition to empirical research, theoretical research is needed to understand and formulate what exactly the outcome of the interviews is.

REFERENCES

[1] Holzblatt, K. (1993). Contextual Inquiry: A Participatory Technique for System Design. In D. Schuler & A. Namioka (Eds.), *Participatory Design: Principles and Practices*. Lawrence Erlbaum Associates.

Integrating the Strengths of Cognitive Emotion Models with Traditional HCI Interaction Analysis Tools

Mark Springett
Computing and Multimedia Technology
Middlesex University, UK
m.springett@mdx.ac.uk

Effie Lai-Chong Law
Department of Computer Science
University of Leicester, UK
elaw@mcs.le.ac.uk

ABSTRACT
Norman's Action Cycle has commonly been applied as an interaction analysis tool in the field of HCI. In wake of the recent shift of emphasis to user experience (UX), the cognition-based Action Cycle is deemed inadequate to explicate affective experiences such as happiness, joy and surprise. Model's based on Appraisal theories, focusing on cognitive accounts of emotion, are more relevant to understanding causes and effects of feelings arising from interacting with digital artefacts. We aim to explore the compatibility between these two genres of model.

Keywords
Cognition; Emotion; Appraisal; Action; Design; User experience; Withdrawal

INTRODUCTION
This paper reports work towards integrating models of emotional factors from the psychology literature with applied models of cognition used in HCI design and evaluation. In particular it analyses compatibility between cognitive accounts of emotion emerging from, among others, the work of Scherer (2002), Ortony et al (1988), Coulson (2004), and established approaches, particularly Norman's (1988) model of display-based, which are used to understand goal-based cognition in interaction and formative evaluation of usability factors, characterising reactions to interaction events, their causes and their effects.

The motivation for this work is to find useful theoretical tools that accommodate both analysis of traditional usability concerns such as comprehensibility of feature cues and feedback, and what are typically referred to as 'experience' factors, where an affective response emanates from encounters with technology. In doing so we hope to better explain the relationship between usability and user experience factors in design.

User experience research does not yet provide fine-grained diagnostic tools capable of pinpointing and understanding elements of designed systems that may undermine positive user experience. Typically UX evaluation tends to deal with overall reactions to the interactive experience. More fine-grained analysis may give designers a better insight for design iteration where a feature or interaction event has had a pivotal effect on user experience or behaviour. In turn this may help designers refine systems at the feature level, and repair what can be termed 'UX bugs' at the interface.

The motivation for this work is that current instruments for valuating user experience do not typically analyse cause and effect in experiential episodes. Typically experience is analysed through retrospective accounts provided by user subjects. These include accounts immediately after an interactive session in which subjects report their summative reactions.

No method currently exists that supports a cognitive account of the emotion through analysis of interactive sequences. The nearest to this is the Sensual Evaluation Instrument (Isbister et al 2006), which stimulates user expression of emotion by interacting with objects. Other approaches try to trace critical incidents by measuring physiological changes in subjects through heart monitors and galvanic skin monitors. However, these provide no more than markers showing where something (in the design or otherwise) affected interaction. What our work aims to provide is a framework for analysing interaction, and linking observed (critical) incidents with antecedents and consequences, to truly understand the role of affect in user reactions to systems. It works on the assumption that a fine-grained causal account of design features' influence on users is required to inform iterative design for optimised user experience.

TRANSFERRING CONTEMPORARY EMOTION THEORY TO COGNITIVE MODELS OF HCI
The exercise reported here is the exploratory integration of cognitive accounts of emotion with theoretical and practical tools for analysing cognition during interaction. The prima facie attraction of this exercise is that the two genres of model have complimentary strengths that can usefully integrate to produce effective user experience evaluation tools. To explain this notion we look in turn at the strengths of each genre.

Strengths of HCI Interaction Models
A key strength of interaction models such as the one described in Norman (1988) is that they facilitate an analysis of causal relationships when applied to interaction events, providing a baseline for understanding antecedents and consequences of system appearance and behaviour. In usability evaluation this contextualises the influence both of prior dispositions (user state of knowledge, background, expertise level etc), as tributaries of user behaviour. It also facilitates the investigation of problem genotypes (root causes of user problems) emanating from error phenotypes

(overt symptoms of a problem detected during interaction). The expression of user and system actions as a connected sequence provides a dynamic mechanism for this. Existing models in the literature add accounts of key catalytic elements in this process. These include accounts of the nature of user mental processing, levels of expertise and experience and the knowledge resources recruited during interaction. Critically, this includes internal and external resources. A prima facie match between the projects of understanding instrumental usability factors in evaluations and affective episodes at the interface lies in this synthesis of internal and external factors. Just as usability problems can frequently be explained with reference to mismatches between the external (the image and behaviour of the machine) and the internal (the users cognitive) resources, a cognitive account of positive and, more critically, negatively valenced encounters can be understood in terms of a similar synthesis of the internal and the external.

Appraisal Theories of Emotion

There are two key elements of the class of emotion theories known as appraisal theories. One is that they reject the conventional taxonomy of distinct emotional states (Ortony et al 1988). Common language tends to embed a naive theory of emotion as falling into distinct categories represented by linguistic tokens such as happy or angry. These are seen as being of little use in understanding the concept (see Boehner et al 2007). The second element is the general believe that the emergent process, the genesis and consequences of emotional arousal are of interest, rather than the qualitative, experienced episode. Emotion is modelled in terms of contextual factors that determine action. The genesis, expression and time-course of emotion arises from a multiplicity of factors or contexts (Coulson 2004). Emotional arousal and the appraisals that may give rise to arousal are distinct and separate in nature. We argue that useful accounts of experience in human computer interaction are a matter of understanding the concept of appraisal and arousal. One crucial aspect is conscious experience of 'emotion'. It is accepted that emotional arousal is felt, experienced and expressible by the individual. Appraisal, however, may not be conscious, it may occur without the individual being explicitly aware of it. This is a key consideration when applied to some phenomena of interest in UX research. Compare a sudden event in a video game for emergency response training to an accumulation of 'concerning' events in a social network encounter. In the former case the sudden onset produces a quick and compelling physiological reaction. By contrast, weakening trust in the identity and integrity of a chat room correspondent could emanate from gradual accumulation of appraisals.

Norman's Model of Action Revisited

In Norman's original model execution has three phases, goal generation, intention forming and translation into a sequence of actions (Norman 1988). The forming of an intention implies generating expectancy of the features that will be encountered. This is characterised as a matching process between internal representations and interface features. These include container metaphors and individual feature representations. A visual scan takes place involving a search for the best match between interface features and the user's goals (Howes & Payne, 1990). The three stages of execution are: perceiving and understanding the state of the world, comparing the state of the world to the intention, and assessing progress towards a goal.

The use of the action cycle as a tool for identifying and characterising usability bugs is established in HCI literature (e.g. Hartson et al., 1999; Springett, 1998). Typical usability problem phenotypes are associated with individual phases in execution specification, physical performance of action, and evaluation. As such these serve as key staples in establishing the 'story' of a critical incident. The establishing of links between phenotype and genotype (root causes), or the tracing of 'critical threads', is key to gaining a deep understanding of usability problems. This trace of critical threads is central to error analysis both in contrived evaluation studies (e.g. think-aloud protocols) and in error studies in the field.

Where a system is 'affect critical' the cycle of action described by Norman (1988) can be seen as a legitimate, but incomplete account of cognition. The account of 'mental actions' has been used in accounts of HCI usability for several types of system. However, it requires a richer explanation of how the mechanics of goal-directed cognition combine with affective reactions to interface phenomena and events.

Events in the context of this analysis could be events occurring as system feedback in response to user action and interface events that isn't directly a response to user action. An event can also be an appraisal as a result of the user scanning a visual image. Therefore we can think of appraisals as occurring at key points in this cycle, including visual scanning in early stages of the execution phase.

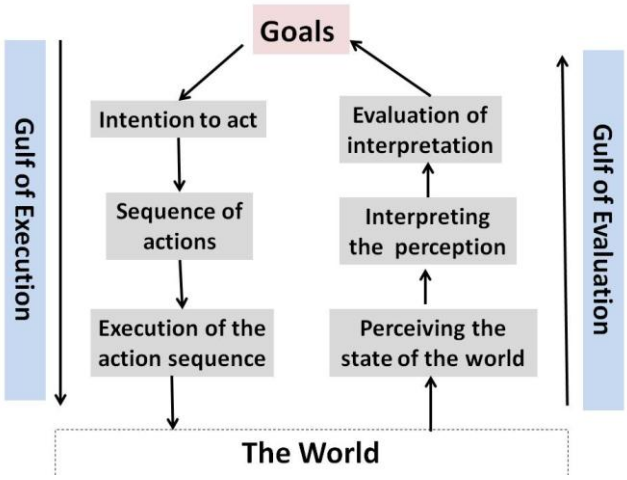


Figure 1: Norman's action cycle (1988)

Stimulus Evaluation Checks

Scherer (1984) proposes taxonomy of ways in which individuals evaluate information and events. These are:

Novelty check: This is a check to see if external or internal stimulation has changed. Internal stimulation could be a triggered memory for a future event (e.g. an appointment). External stimulation may include a match between expectations of system behaviour and new system behaviour.

Intrinsic pleasantness check: This detects a positive or negative valence, determining approach behaviour or withdrawal/avoidance.

Goal/need significance: This is composed of evaluations of relevance, expectation, conduciveness and urgency. Assessment of relevance relates to the selection of features in action execution, and match to goals in the evaluation phase of Norman's model. Expectation and conduciveness equally seem to express the phases of interpreting and matching to goals expressed in Norman's model.

Coping potential: This evaluates causality, the level of control the individual has over its consequences, and the ability to adapt to cope with it.

Norm/self compatibility check: This involves normative judgments about the event. This may be a match between an internal standard or norm. In e-service use for example it may be a comparison of system design of behaviour to expectations of service or quality of design. It also has a socio-cultural dimension where the norms of others and accepted cultural norms are brought to bear.

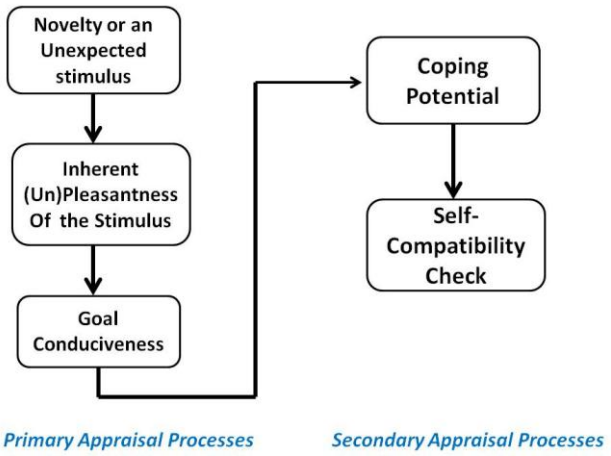


Figure 2: Overview of Scherer's Appraisal Model

Example – E-Commerce Trust Propagation

In this example, we consider the influence of the novelty, intrinsic pleasantness and norm/self compatibility checks. Several studies (Riegelsberger et al, 2005; French et al, 2006; Sillence et al 2007) suggest that display factors have a significant influence on trust-related judgements. E-Commerce encounters involve the perception of signs (interface appearance) and events that may ether positively or negatively reinforce trust in the competence of the system and the identity/integrity of the organisation it represents. Trust propagation in e-commerce is seen as a journey from initial expectations of the organisation and encounter, through the first encounter with the website and the completion of service transactions (French et al 2006). Critical phases in which the e-customer's relation is mediated through interface features and behaviour, including overt tangible trust signs, and sundry aspects of the interactive session that could potentially affect attitudes and behaviour.

The match between expectations and what is encountered (novelty check) may be pleasing revelation of positive trust re-enforcers such as trust seals or third-party associations. The 'warmth' of this re-assurance (intrinsic pleasantness) fortifies the relationship between individual and organisation. However, this could also emanate from aesthetic factors such as a pleasing colour scheme or familiar cultural references. In the negative case an event that infuriates, such as the deletion of input data, or unexpected task steps, may confound positive expectations of the organisation. This may also include the norm/self compatibility check, where the user compares the demands made by the system to their general perceptions of what is reasonable. Similarly, requests for personal information may cause a negative reaction when compared to culturally-mediated perceptions of the limits to invasion of personal privacy.

A Reinterpretation of the Basic Model of Action

Below we revisit key phases in Norman's model of action, adding concepts referred to in the theories considered above.

Goal formation: Goal formation implies the generation of satisfaction criteria. These could be criteria such as safety/security that are not explicitly part of the task model.

Intention Forming: Implicitly involves expectations of system features and behaviour.

Scan matching feature/operation (Appraise image): The scan of the interface to find features must simultaneously imply appraisals that assess match with expectations, opportunities for action, and also assessment of 'warmth/hostility' and other terms often referred to in UX taxonomy. Positive valance emanates from detection of such positive qualities and negative from those suggesting threat, disturbance or disappointment. The former is likely to reinforce **approach** behaviour the latter **withdrawal**, dependency on the strength. A slight concern that the system image isn't conveying honesty or reliability may not itself be sufficient to cause withdrawal, but may be an input into appraisal of future events.

Perceive feedback/ primary appraisal: At this level of immediacy, primal cognitive functions are likely to be most influential, whether sudden and high impact evaluation (e.g.

a shock reaction such as a loud noise) or a low impact evaluation (e.g. a transient awkwardness on completing a manipulation).

Understand/interpret/appraise change: Assessment and appraisal of the event is linked to Norman's concept of understanding and interpreting feedback from the system as a result of user action. Again there may be affect with significant force that causes withdrawal (perhaps abandonment) or simply a re-evaluation of approach and the necessary conditions for continued action.

Match to current/overall goals: In strict terms the satisfaction of a goal is the completion of a recognised sequence of task-steps. However, if experience factors are an additional feature of this account, then it can be argued that this extends to a wider consideration of the overall conditions for proceeding with goal-directed action. From the 'pure' usability standpoint goals may be supported, as progress towards them is satisfactory supported through action cures and feedback. However, appraisals potentially lead to re-evaluation of user motivation and acceptance of system. If a sequence of appraisals, for example, has the effect of reducing trust in the system and those perceived as being personified by it, the likelihood of withdrawal increases.

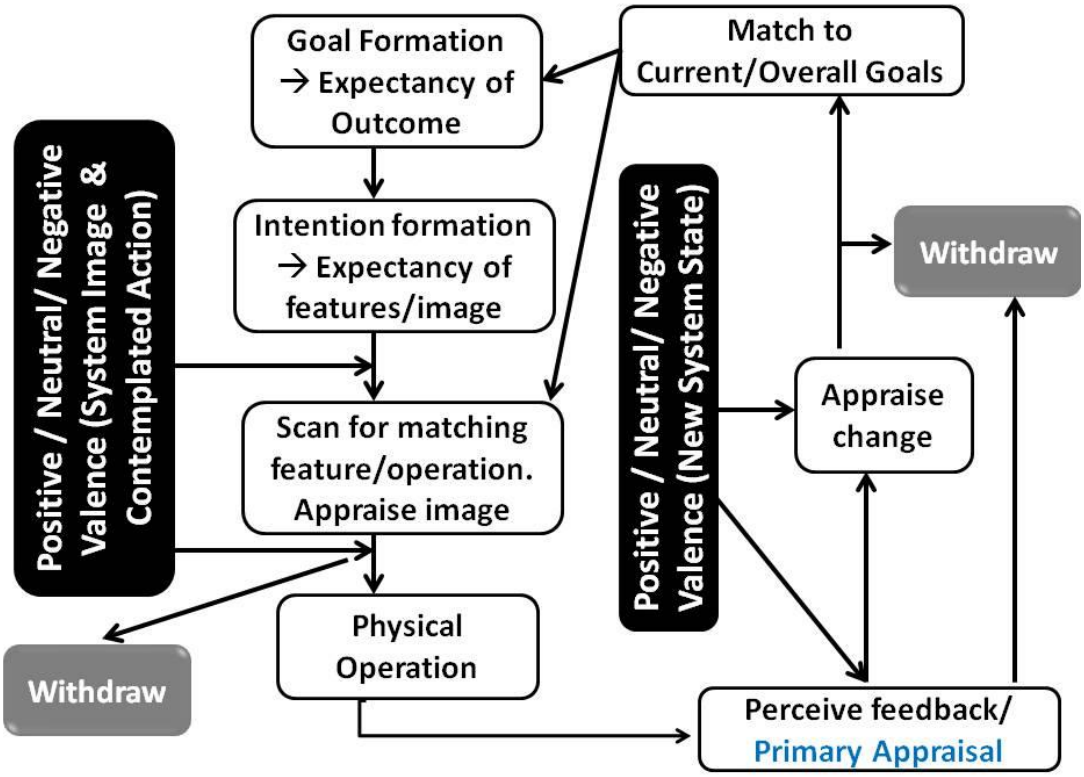


Figure 3: Integrated models of Norman's action cycle and Scherer's appraisal model

SUMMARY:EMOTION OR APPRAISAL-BASED ACTION?

The indications both from literature studies and from the application of merged models to affect-critical systems are that emotion is something of a redundant notion in studies of experience within interaction. What is of greater interest is the series of cognitive appraisals that are applied to phenomena and events during interaction and the consequences that this has in terms of user behaviour and summative evaluation of experience. Norman's action model was a baseline description of action which analysts and researchers could apply to assess gulfs of execution and evaluation (Figure 1) in usage of a number of products. Likewise a model that combines the key elements of this model with accounts of appraisal provide a baseline for understanding affect in the context of goal-directed user action. Immediate factors such as positive or negative valance and approach or withdrawal (if the stimulus has high intensity) are accounted for within the cycle of task-action. Also, accounts of learning by exploration and synthesis of examples accommodates key appraisals with less high intensity that contributes to a relatively slow affective onset.

The six contexts described in Coulson (2004) (i.e. Event, Agent, Interpersonal, Topographical, Historical and Embodied) emphasise the factors that become particularly relevant dependent on the type of design problem considered. For example, the interpersonal context explains

appraisals in which the intentionality of e-commerce organisations is deconstructed and interpreted through encounters at the interface. The same context characterises the sense of self that emanates from assumed characters in game play.

Formative design and evaluation benefit from having runnable models that can be used, either in the form of an explicit procedure, or as a tool for thought. Theoretical tools that integrate actions of display-based cognition and appraisal can analyse both the pragmatic aspects of usability and the affective factors that influence user behaviour and judgement.

The integration of Norman's theory of action with constructs from appraisal theories has the potential to produce useful and usable tools for understanding user experience factors during interaction. Questions relating to the true nature of the relationship between usability and user experience remain, but there is clearly value in understanding these factors in an integrated way. Future research can usefully be directed towards developing analysis tools that can facilitate the application of this in design and evaluation.

REFERENCES

1. Boehner,K. , Depaula,R., Dourish, P., and Sengers,P., How emotion is made and measured Int. J. Human-Computer Studies 65 (2007) 275–291
2. Coulson, M., (2004) An 'Everything But' Framework for Modelling Emotion, in Proc, AAAI Spring Symposium on "Architectures for Emotion"
3. French, T, Springett, M., Liu, K., (2006) 'Towards an E-Service Semiotic Trust Framework',Procs. 4th Int. ALOIS Conference, Boras Sweden,175-196.
4. Hartson,H. R., Andre,T.S., Williges, R.C., van Rens, L.,: The User Action Framework: A Theory-Based Foundation for Inspection and Classification of Usability Problems. HCI (1) 1999: 1058-1062
5. Howes, A. & Payne, S.J. (1990). Semantic analysis during exploratory learning. In J.C. Chew & J. Whiteside (Eds.) *Empowering People - CHI'90 Conference Proceedings: Human Factors in Computing Systems (Special Issue of the SIGCHI bulletin).* Seattle, WA: ACM. 399-406.
6. Isbister, K., Hook, K., Sharp, M., and Laaksolahti, J. (2006). The sensual evaluation instrument: developing an affective evaluation tool. In Proc CHI '06. ACM, New York, NY, 1163-1172.
7. Norman,D.A., (1988), *The Design of Everyday Things*, Basic Books
8. Ortony, A., Clore, G., Collins, A. 1988. *The cognitive structure of emotions.* Cambridge: Cambridge University Press.
9. Riegelsberger, J., Sasse, M.A., and McCarthy, J.D., The mechanics of trust: A framework for research and design: *International Journal of Human-Computer Studies 62* (3), 381-422
10. Scherer, K.R. (1984). On the nature and function of emotion: a component process approach. In K.R. Scherer & P. Ekman (Eds.), Approaches to emotion, (pp. 293-317). Hillsdale, NJ: Erlbaum.
11. Scherer, K. R. (2001). Appraisal considered as a process of multi-level sequential checking. (Scherer, K. R., Schorr, A., Johnstone, T., Ed.).Appraisal processes in emotion: Theory, Methods, Research. 92-120.
12. Sillence, E**.,** Briggs, P. Harris, P, Fishwick, L. (2007). How do patients evaluate and make use of online health information? Social Science & Medicine, Volume 64, Issue 9, Pages 1853-1862
13. Springett, M., Linking surface error characteristics to root problems in user-based evaluation studies. Proc. AVI 1998: ACM Press 102-113

Some Reflections on Evaluating Users' Experience

Alistair Sutcliffe
Manchester Business School
University of Manchester
Booth Street West
Manchester M15 6BP, UK
ags@man.ac.uk

Jennefer Hart
Manchester Business School
University of Manchester
Booth Street West
Manchester M15 6BP, UK
jennfer.hart@postgrad.mbs.ac.uk

ABSTRACT
We report experience in using a variety of techniques to evaluate user experience, ranging from interviews to observation and surveys. UX was evaluated in sessions by an experimental comparison of three different websites, and over a longer six-month period by a combination of diary studies, interviews and surveys. The advantages and disadvantages of each method are discussed, as well as feedback on these techniques from industrial interaction designers and UX practitioners.

Author Keywords
User Experience, Evaluation, Longitudinal studies, Experiments

INTRODUCTION
In this position paper we share our experiences in evaluating user experience (UX). While these reflections are based on academic research we believe they are relevant to industrial evaluation of UX; furthermore, we have presented out findings to UX practitioners and the feedback gained from these interactions is also reported.

Unlike usability, which can have reasonably objective criteria in terms of observed user problems, user reports of difficulties, etc., UX is a more subjective set of qualities which we consider form part of design quality judgement. Components of user experience may vary from assessments of utility and functional worth, closer to effectiveness in classic usability definitions [8], to content and brand in products, aesthetics as well as interaction-related concepts [5].

To evaluate UX, new techniques are required to assess the user's perception of presence, attractiveness of the media, flow and engagement in interaction as well as excitement and affect consequent on interaction [7]. UX also needs to be evaluated over time, since early experience may be determined by aesthetic perceptions, whereas later experience is conditioned by interaction, usability and utility [10].

Our evaluation approach is based on triangulation of evidence, combining observation of users during the interactive experience, followed by post-test questionnaires and debriefing interviews to follow up on any critical incidents observed during interaction. Debriefing interviews may also include a free recall memory test, asking users to list the first few items about their experience which come to mind, which may be memory of features, common activities, general aspects of the design, feelings and emotions or problems encountered during interaction. Memory reports can then be probed to explore general opinions and affective responses to particular product features. Concurrent protocols or question-led interventions frequently recommended for usability evaluation [11] are not appropriate since they would disrupt flow and presence. Users are requested to complete a set of tasks or test the application while being observed or video recorded.

We report our experience in experimental evaluations and a longitudinal study of UX with medical students using iPads for e-learning over six months. We have used a variety of methods including observation of use during collaborative workshops, diaries recording experience vignettes on a weekly basis, interviews to follow up diary studies at three check points, and questionnaires to supplement qualitative data to evaluate UX, utility, usage and usability at the start, mid- and end-point of the study. In the following section we reflect on the variety of experimental and longitudinal studies we have conducted that employed several different methods of assessing UX.

CONTEXT
Since, we are not reporting a case study, we will give only brief details on the background of our studies. The experiments aimed to compare three websites in the same domain, online art galleries (National Gallery, Louvre and Google Art Project); these have varying degrees of sophistication in interaction design, to investigate user experience over repeated exposures. After completing the pre-test demographic questionnaire, participants were shown screen shots of each of the three website homepages for 0.5s (initial exposure), followed by two brief questionnaires (affect and website quality [6]). After a short gap users completed task 1 (general navigation), then they completed the affect questionnaire, followed by the second task (detailed exploration and interaction), and four questionnaires (affect, website quality, immersion/presence

and overall preference). A semi-structured interview to elicit participants' preferences and experiences while interacting with the websites completed the study. Participants were a mix of post-graduate students and university staff; further details of the material and methods can be found in [3].

The longitudinal study, which is nearing completion, was closer to a case study in format. Participants were 51 third year medical students in the University of Manchester, age range 24-29, approximately 65% female to 35% male, who were in their first year of clinical practice in hospitals in Manchester and Preston, UK. The context of the study was their use of iPads provided at no cost by the University as an e-learning collaboration resource. The iPads were intended for access to Internet resources such as *British Medical Journal* libraries and a variety of specialist education medical applications, as well as use of Blackboard, the University's collaborative learning environment. All the participants were novice users of iPads and tablet technology, although most (75%) had iPhones. Data was collected by diaries which requested participants to enter a weekly assessment of their activities (work and social), and experience both quantitatively on a brief scale (enjoyable, boring, frustrating, useful, engaging), and qualitatively by reporting the best and worst experiences they remembered for four specific activities. Diaries were supplemented by questionnaires on activities and experience pre-study (month 0) for expectations, then at months 3 and 6 to record experience on four scales (activities, aesthetics, affect and satisfaction/usefullness). Interviews at months 0, 3 and 6 probed for best and worst experiences, critical incidents if any, overall assessment of their iPad for work and leisure use, with reasons for adopting or possibly rejecting the product. Our objective was to track their activities with the device and different applications over a 6-month period, evaluate the user experience and investigate reasons for adoption or non-adoption of the technology.

QUESTIONNAIRES

Since UX is often considered to be an attitude or felt experience, questionnaires are a natural way to capture users' opinions. We have used established scales from the literature; for example, aesthetics can be evaluated using either Tractinsky's expressive aesthetics inventory [9] or Hassenzahl's hedonic scale [6]. The classic aesthetics scale overlaps with traditional usability, so it could be omitted. However, the pleasure inventory [9] does provide a useful summative assessment of satisfaction and pleasure with the product, while for information quality the Bernier scale evaluates content and service quality and assesses overall utility and satisfaction. In other studies we have combined Hassenzahl's AttrakDif for hedonics and pragmatics (usability/utility) with affect measures [2,13]. However, our motivation for using existing scales is to reduce the criticism from academic reviewers when new self-designed scales are proposed. In practice we have found that most scales pose problems in interpretation by users and may produce inconsistent results, where reported attitudes are extremely susceptible to framing effects. In other words, the answers depend on the prompt, scenario, task and domain the user is given in the evaluation exercise [4,5]. When capturing immediate post-test experience, the time taken to complete questionnaires poses further problems: by the time the users report their attitude their memory may have waned and the mood/affect could be different. Simple bipolar questions and short scales (<10 questions) help to capture elements close to UX that are prone to waning effects such as flow and presence. Questionnaires are therefore best used with some caution, and care should be taken over the framing effects, which may colour the user's opinion. Indeed we have demonstrated that users' ratings of the same product post experience will change radically from prompting with different post-test scenarios of use [15]. A further problem arises when quantitative attitude ratings are compared with qualitative data gathered from interviews after experience on the same product by the same users. While overall judgement is broadly consistent, considerable differences do arise between opinions derived from scales and textual analysis of recorded interviews.

INTERVIEWS

We use post-test interviews to understand users' understanding of their experiences. First we use immediate memory free recall tests; these report the users' first 2-3 'most remembered' features/aspects of the product/website, , then rate them as a good/bad/neutral experience. Participants struggle to remember after the first two items and often need further prompting, possibly because the act of verbalising the first item hinders recall of subsequent items. Memory recall captures design features and general opinions which are salient in the user's memory immediately after interaction. If the experience was vivid the chances are the design sticks in the user's memory; conversely, if the experience was awful, usability problems are also reported. We then conducted a semi-structured interview asking the user to recollect, by revisiting the application, highlights and downsides of their experience, reporting the design features involved and the reasons why they liked or disliked them. When interviews are covering a period of experience (i.e. a week), we first ask the users to report the main activities for which they use the product/application, as a means of situating further questions about experience and usability problems. Interview data is analysed with a coding categorisation methodology used for academic research; however, the results are immediately accessible from listening to the interview recordings.

We speculate that the difference between our measures of UX arising from questionnaires and interviews arises from

the different reporting methods. Questionnaires capture general opinion after experience, whereas interviews encourage more reflection on the design and experience, and users' opinions may well change during the reflective process. This has important implications for UX research as well as for practical evaluation. User judgement may shift during post-experience reflection when different trade-offs are considered rather than a more 'gestalt' affective response to questionnaires.

OBSERVATION

Observation during group settings is more difficult given the number of users. We were interested in interaction, manifestation of affect that might indicate experience, as well as patterns of use, i.e., how the iPad functioned as a shared collaborative device or as a solo application. For affect we attempted to record posture, gesture, facial expression and voice prosody as affective indicators of user experience. For instance, a hunched posture focusing on the interface suggests engagement, whereas lack of attention and a relaxed posture suggest the opposite. Facial expression and voice prosody are good indicators of affect, indicating pleasure, joy and surprise as positive reactions to experience or frustration, anger and anxiety as negative aspects. Observing affect and experience in several users concurrently in a group was not feasible; since one observer can only focus on one or two key aspects at once; otherwise, sessions have to be video recorded for subsequent analysis, a time-consuming process even with informal inspection-based analysis. We also noted a further confound, that affect response might be part of a conversation about the domain subject matter or reaction to inter-personal dialogue. It was difficult to differentiate product-related reactions from social interactions. While observing individual users might simplify this problem by eliminating social responses, the problem of disentangling product interaction-related affect from responses to content remains.

MEASURING EXPERIENCE OVER TIME

We used two strategies: first a repeated exposure experiment to assess experience in the short term, i.e. a repeated task within one session; then we used a diary technique in a longitudinal study of multi-session UX.

Experimental Evaluation

We conducted a repeated exposure experiment in which users carried out slight variations of the same task twice to evaluate how their attitude changed with interaction. Users first viewed the home page of a website to evaluate perceptual aesthetics prior to interactions, then they carried out the set task in two successive sessions with a small (<5mins) gap during which time they completed a feedback questionnaire. To minimise the interruption of the 'flow' experience during the experiment, only a short 9-item affect scale was completed after the first task, then the complete set of questions (aesthetics, usability, flow and presence) was completed after the second task. The slight variations between the tasks were intended to counteract waning in interest and to introduce users to new interactive features.

The results demonstrated a dramatic effect with all UX measures (aesthetics, presence, usability, affect) increasing from initial exposure to task one, with a modest increase between task one and task two. We tested three websites in the museum domain with different interactive features, showing that the designs with more interactive features (3D fly-through interaction, avatars, active objects and graphical objects) had more positive UX ratings than the baseline site with standard menu-link navigation [3].

Our experiment did show that UX can be evaluated over time; however, there was some evidence of waning effects even after two task exposures. Asking users to repeat the same task several times might have produced more interesting experimental data, but in terms of ecological validity, this is limited. Business applications that involve form-filling data entry and repeated tasks will focus UX on usability, effectiveness (speed of operation) and utility. Other applications where the task is unlikely to be repeated in successive sessions are more difficult to evaluate with a controlled experiment, hence we adopted diary studies to track the diversity of use and experience over time.

Diary Studies

We used a structured format to encourage users to report on specific aspects of their experience, with instructions to recall episodes of use both good and bad in the previous week. The accuracy of completion varied between participants; a few were verbose and consistent in reporting their experience, but most were terse and consistency declined as the study period progressed. Good completion rates and participation were maintained in the six-month study, although continual reminders and incentives at the end of the period were necessary. Negative experiences, i.e. usability problems, may be reported more frequently than positive experience, so diary methods can introduce a bias since human memory tends to be more vivid for unpleasant rather than pleasant experience [1,12]. Reporting of activity was more accurate and consistent compared to attitude data, where most positive experience reports tended to be general impressions, e.g. 'fun', 'enjoyable', 'nice design'.

Questionnaires integrated with the diary were completed more consistently although many reminders had to be sent to improve completion rates. Diary studies have proven worthwhile for capturing activity data over time: what users are doing with their device/applications; however, how they feel is not captured so accurately in the qualitative record. Accordingly we have included a brief questionnaire to capture affect as part of the weekly diary completion schedule. Periodic questionnaires using cut-down versions of the affect and experience scales were completed reasonably consistently; however, qualitative data in diary reports was less useful for tracking experience over time.

The connection between experience and the report thereof in diaries is difficult to assess. Negative reports did provide insight into usability problems but it was more difficult to track their resolution. When the problem was not reported in subsequent weeks, this may indicate the users had discovered a work-around, or that the problem might persist but was simply not worth reporting. In contrast, positive experience was more difficult to assess from qualitative reports. Questionnaires were periodic (monthly), so the attitude captured may reflect a general impression over that period or, more likely, recent memory over the last few days, since human episodic memory tends to be biased towards salient, unpleasant events and recent experience. Diary studies prompt for day-by-day recall of experience can be employed to try and counteract this problem.

PRACTITIONER FEEDBACK

We presented our methods and results of the studies to meetings of the Northern User Experience meetings and discussed them in a workshop at the NUX conference in November 2012. The NUX group is composed of industrial practitioners: usability and user experience professionals, consultants and interaction designers in the North of the UK. The result of the studies provoked considerable interest in terms of design implications for interactive features: the relative importance of content and functionality versus aesthetics and design in overall user preference and satisfaction. Reaction to the methods focused on interviews and diary studies being the more practical, while questionnaires and observation were too time-consuming to be cost-effective in commercial practice. While interviews were already part of their normal practice for evaluation, diary studies were a novel approach which several members were interested in adopting with their clients, not only for evaluation purposes but also as a means of fostering the customer relationship.

CONCLUSIONS

A combination of techniques for evaluating UX has proven effective for our research, with experiments providing insight into the different experience between products with more or fewer interactive features and changes in experience over short time scales. However, even during the experiments we found differences in participants' assessment of the same products in quantitative and qualitative data. Furthermore, general measures of experience masked individual differences and how groups of users reacted to products. Analysis of data at the individual level suggested that users may have styles that shape their experience, such as enthusiasts who rate products highly and make no criticisms, compared to adopters who rate products less well and do make criticism such as noting usability problems. Finally there are non-adopters who have little positive comments to say about a product. Individual styles may therefore colour evaluation of user experience. Some users could form 'gestalt'-style opinions of products which then influence their reported experience both quantitatively and qualitatively, while other users may be more reflective and come closer to an evidence-based evaluation of experience.

In the longer term, evaluating data analysis is still in progress so we have focused on our experience of the data capture evaluation techniques. None are ideal: interviews are time consuming and difficult to schedule consistently over a 6-month period; periodic questionnaires are reasonably effective at capturing affect and UX, but only supply a limited number of data points over time; finally, diary studies have been effective in capturing activity but are less effective in providing evaluation of positive user experience. For practical evaluation a combination of the four techniques we used may be too expensive so we would recommend interview with periodic lightweight questionnaires as the best means of assessing UX in the long run.

There are other UX evaluation methods which we did not use in our studies, notably physiological measures such as GSR (Galvanic skin response), heat rate, EEG (Electro-encephalogram), etc. While these measures have the merit of objectivity, their downside is intrusive instrumentation which could disrupt user experience, particularly for aspects related to flow. Further problems which dissuaded us from adopting these measures were the cost, and relating the results to user experience; for instance, GSR and heart rate are good measures of arousal but are hard to interpret in terms of flow, engagement, presence and other facets of UX.

We do not propose a UX evaluation framework; instead, the following informal guidelines provide some advice based on our experience:

- Calibrate evaluations according to the domain: for work/goal-oriented applications, content, functionality, and usability may make more important contributions to overall experience than interaction, aesthetics and user engagement; whereas for entertainment applications the opposite is likely with aesthetic, interactivity/engagement then content and usability.
- For evaluating UX in the short (within session) time scale, use measures of interaction/flow and immersion/presence to assess reaction to interaction design, with emotional arousal and mood for an overall evaluation of users' feelings.
- For evaluation of UX over longer durations, measures of activity (self-reported or automatically logged) with attitude reports via diary studies and mini-questionnaires are more effective.
- Where possible use mixed methods: quantitative ratings from surveys with qualitative interpretation of interviews to discover the reasons for user attitudes and preferences. The remaining three guidelines give hints

on eliciting experience measures which can be applied to interviews or observations.

- Emotional responses can be detected from facial expressions, gestures and body postures, possibly supplemented with verbalisations if made. Free recall memory of feelings is a useful cross-check to questionnaire replies.
- Flow-interaction: observation of the pace of interaction should focus on any breaks in flow, e.g. looking for critical incidents, hesitations and signs of frustration and annoyance.
- Immersion-presence can be judged from users' attention to the interface and from flow. Any disruption to flow is also likely to degrade presence. In memory recall, reference to design features, especially interface controls, indicates poor presence; in contrast, general impressions and feelings about involvement indicate good presence.

The above guidelines can be seen in the resources perspective of evaluation practice where studies of industrial practice have demonstrated that methods tend to be composed flexibly by practitioners within the constraints of cost, the domain, skills available and any other contextual factors [16,17]. The first four guidelines provide advice on technique-domain tailoring, while the latter three propose measurement instruments to capture specific aspects of user experience. Customisation and adaptation of methods in practice has also been a long-established finding in software engineering where the concept of 'method engineering' [14] has developed to investigate flexible composition of procedures and techniques in different application contexts.

A further resource reflection concerns the cost effectiveness of techniques and the evaluation goals. While most of our studies were driven by academic goals to understand UX at a theoretical level, most evaluation is driven by formative concerns of design improvement. This will focus on the aesthetics and interaction design, which may be better investigated by interviews and observation, since questionnaires rarely capture data at the feature level.

REFERENCES

1. Baddeley, A.D. (1986) *Working memory*. Oxford: Oxford University Press.
2. Berlyne, D.E. (1960) *Conflict, arousal, and curiosity*. New York: McGraw-Hill.
3. Hart, J. and Sutcliffe, A.G. (2013) Love it or hate it! The UX of interactivity and user types. To appear in *Proceedings CHI-13*. New York: ACM Press.
4. Hartman, J., Sutcliffe, A.G. and De Angeli, A. (2008) Towards a theory of user judgment of aesthetics and user interface quality, *TOCHI*, 15(4),1-30.
5. Hartmann, J., Sutcliffe, A.G., and De Angeli, A. (2007) Investigating attractiveness in web user interfaces. *Proceedings CHI-07*. New York: ACM Press.
6. Hassenzahl, M. (2004) The interplay of beauty, goodness, and usability in interactive products, *Human-Computer-Interaction*, 19(4), 319-349.
7. Hassenzahl, M. and Tractinsky, N. (1996) User experience: A research agenda, *Behaviour & Information Technology*, 25(2), 91-97.
8. ISO (1998) ISO 9241-11: Ergonomic requirements for office work with visual display terminals (VDTs), Part 11: Guidance on usability. ISO.
9. Lavie, T. and Tractinsky, N (2004) Assessing dimensions of perceived visual aesthetics of web sites, *International Journal of Human Computer Interaction*, 60, 269-298.
10. Lindgaard, G., Dudek, C., Sen, D., Sumegi, L. and Noonan, P. (2011). An exploration of relations between visual appeal, trustworthiness and perceived usability of homepages. *TOCHI*, 18(1), 1-30.
11. Monk, A., Wright, P., Haber, J. and Davenport, L. (1993). *Improving your human-computer interface: a practical technique*. London: Prentice Hall.
12. Neisser, U. (1976) *Cognition and reality*. San Francisco: W.H. Freeman.
13. O'Brien, H.L. (2010) The influence of hedonic and utilitarian motivations on user engagement: The case of online shopping experiences. *Interacting with Computers*, 22(5), 344–352.
14. Rolland, C. (2009) Method engineering: towards methods as services. *Software Process: Improvement and Practice*, 14(3), 143-164.
15. Sutcliffe, A.G. and De Angeli, A. (2005) Assessing interaction styles in web user interfaces. *Proceedings Interact 2005*. Berlin: Springer.
16. Whitefield, A. and Sutcliffe, A.G. (1992) A case study in human factors evaluation. *Journal of Information Systems and Technology*, 34(7), 443-451.
17 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. *International Journal of Human-Computer Interaction*, 27(10), 940-70.

Hidden Biases in Semiotic Engineering Introducing Communicability Evaluation to Multi-Touch Interface Design

Jan Derboven

CUO | Social Spaces, KU Leuven-iMinds
Parkstraat 45 bus 3605
jan.derboven@soc.kuleuven.be

ABSTRACT
This paper describes the use of Semiotic Engineering methods in the context of multi-touch interaction. Focusing on the analysis of user testing data, we describe the transferability of Semiotic Engineering methods (specifically the Communicability Evaluation Method) across two interaction paradigms. While Communicability Evaluation is traditionally used for the analysis of graphical user interfaces (GUIs), we assess its applicability in multi-touch interaction through a case study.

ANALYZING 'NEW' INTERFACES WITH 'OLD' METHOD-RESOURCES

While multi-touch technology has been around for a few decades, the increased interest in this technology since the early 2000s has led to a lot of new research in interaction design for multi-touch applications. Commonly considered a 'new' interface paradigm, multi-touch technology continues to offer numerous challenges to optimize design for interaction using touch and gestures. The case study described in this paper gives an overview of the efforts we made to tackle multi-touch interaction design challenges using Semiotic Engineering methods. We first sketch the original application context of Semiotic Engineering methods. Afterwards, we describe the rationale for applying these methods to multi-touch interface design, and the experience we had with the process of the transfer to multi-touch.

Original Application Context

The foundations of De Souza's Semiotic Engineering theory date back to 1993, with a paper on 'The semiotic engineering of user interface languages' [3]. In it, De Souza frames her work as 'an attempt to give theoretical support to the elaboration of user interface languages', and to 'sketch the basis of a theoretic approach to user interface language design'. Since 1993, this approach has evolved from a basic theoretical sketch to a complete semiotic ontology of HCI, with its own distinct methods. The main novelty in the Semiotic Engineering approach to HCI is the focus on HCI as a specific type of designer-user communication: the actual communication 'is between designer and person, where the technology is the medium' [6], and not between user and system. The strength of Semiotic Engineering lies in this shift in perspective: it can generate a 'new account of known problems' [5], opening up new solutions, and possibilities for design and redesign.

The two main Semiotic Engineering methods, '*epistemic tools* that should not be used to give directly answers to design problems, but to increase the problem solver's understanding of problems and alternative solutions' [5], are the **Semiotic Inspection Method (SIM)** and the **Communicability Evaluation Method (CEM)**:

- **SIM** allows the analyst-expert to analyze the way a specific system communicates the designer's messages to the user. As such, SIM is an inspection method that aims at reconstructing the designers' communication: it evaluates whether or not designers' intent has been communicated effectively through the design choices and interactive content in the system.
- **CEM** focuses on the designers' communication through user observation. Users' interactions with a system are analyzed within a very specific analytic scope, resulting in a reconstruction of the actual designer-user communication that unfolds as the users interact with the system [5].

These methods can either be used on their own, or in combination with each other. The case study described in this paper, however, focuses on the transfer of CEM to multi-touch applications.

CEM is a specific qualitative method that analyses user interfaces based on a semiotic interpretation of user test sessions [5]. CEM primarily focuses on interaction problems, and enables researchers to identify and classify user problems in a fine-grained analysis that adds important nuances to the analysis that would otherwise risk being overlooked. For instance, user problems due to insufficient system feedback are broken down further by CEM into several subcategories, such as users misinterpreting the design solution presented by the interface, users completely unable to make sense of interface icons, etc. In this way, CEM presents an important added value compared to other evaluation methods, in that CEM's detailed assessment of

user problems provides designers with a wealth of detailed information. This information can help in making informed decisions during an application redesign.

CEM offers a semiotic framework to process lower-level observations, and draw more generalised conclusions in order to determine an overall semiotic application profile of an application. This process, from low-level observations to high-level semiotic profile, consists of three stages: *tagging*, *interpretation* and *semiotic profiling*.

The first stage in the Communicability Evaluation Method is to *tag all problems users encounter* according to a predefined coding scheme based on semiotic principles. In order to identify all communicative breakdowns between the user and the designer (i.e., instances where users fail to receive the communication as it was intended by the designers), all test participant actions in a user test need to be recorded and tagged according to the 13 tags proposed by De Souza and Leitão [5]. These tags are expressions that users might utter when being faced with problems during the test[1].

In the second, *interpretative stage*, the goal is to search for the higher-level problems users have in understanding the designers' messages. In order to find these higher-order problems, all *tags are analyzed and organized* according to a number of different perspectives, such as the frequency and context of occurrence of each tag, and the existence of patterns in the sequences of tags [5]. CEM prescribes that the low-level tags should be grouped in a higher-level taxonomy of interaction and usability issues. While CEM proposes its own classification, the specific taxonomy used is left open [9] to other possibilities, such as the usability heuristics by Nielsen (see also [4]) or Shneiderman, or other, more domain-specific heuristics or guidelines.

In the final *semiotic profiling stage*, the goal is an in-depth characterization of the way the interactive product communicates with its users. Based on the previous stages, semiotic profiling can answer high-level questions on the way users interact and communicate with the computer system.

Although they have been applied primarily to point-and-click graphical user interfaces (GUIs) - examples include [8, 10] - the method descriptions offered by De Souza et al. (tools to increase the analysts' understanding of user interface problems) suggest that Semiotic Engineering can be an invaluable theory in understanding and further developing innovative interface 'languages' or paradigms, such as multi-touch interface design.

New Context: A Lot of Multi-Touch Experimentation, but Few Standards

We considered multi-touch interfaces as being particularly interesting for a semiotic analysis, since they are often considered to be 'Natural User Interfaces' (NUIs) [15, 16]. The NUI paradigm is a user interface paradigm that goes beyond the point-and-click, metaphor- based interfaces of GUIs. The objective of NUIs, as described by its advocates, is to deliver intuitive, seamless and 'unmediated' experiences that unfold through natural human input [14].

The above characterization of NUI systems means one the one hand that the indirect input of graphical user interfaces using the WIMP paradigm are replaced by speech, touch and gesture— specifically touch and gesture in the case of multi-touch interfaces. On the other hand, the claim of 'unmediated' interaction seemed to imply that the content itself would serve as the interface: users interact in a direct way with the content, instead of needing a GUI with metaphors and icons to access the content [16]. However, the 'unmediated interaction' claim does not remain unchallenged: Saffer [11] argues that 'metaphor will always play a role with our devices— it is impossible to use or understand them otherwise.' In the same spirit, Microsoft detracts from their 'unmediatedness' claim by proposing that the NUI's 'icons' are representations 'in which a portion of the object stands for the object itself' [16]. This phrase in itself describes a kind of mediation: it can be seen as a short, simple definition of metonymy.

Approaching NUI interaction as a distinct kind of mediated interaction, NUIs can be evaluated using Semiotic Engineering. The interface, then, is seen as a communication channel between designer and user. We see several NUI design issues in which such semiotic analysis can be helpful. One of these issues is determining to what extent users can make sense of the system's interface, and assess the need for explicit user assistance. User guidance, especially in multi-touch environments, is often regarded as unnecessary and merely a quick fix for a poorly designed application [13]. While exploring a system, users should be able to find out which functionality is available with as little user assistance as possible. This view is complementary to the idea that multi-touch interfaces should be 'natural', self-explanatory and intuitive. However, this idealistic view often contrasts with the way users actually explore and use multi-touch interfaces [14]. To learn more about issues such as this user guidance contradiction, we decided to use Semiotic Engineering, as a theory that is potentially well-positioned to shed a new light the way NUI interfaces communicate their functionality to the users.

[1] Each tag in the method targets a specific interaction problem a user might encounter. Examples include *Why doesn't it* ('the user expects some sort of outcome but does not achieve it. The subsequent scenario is that he then insists on the same path, as if he were so sure that some function should do what he expects that he simply cannot accept the fact that it doesn't') and *I can't do it this way:* 'Sometimes the user follows some path of action and then realizes that it's not leading him where he expected. He then cancels the sequence of actions and chooses a different path.'[9].

Stories of Transfer: Mixed Feelings

When we were evaluating MuTable, the multi-touch tabletop project we use as a case study, in late 2009/early 2010, not many guidelines or resources were available for designing good multi-touch user interfaces, or for evaluating them. In fact, the field of multi-touch interaction still is quite new, with much room for additional research: the lack of a clear gestural UI language is brought forward by several authors [7, 15], Don Norman and Jakob Nielsen stating that 'these interaction styles are still in their infancy, so it is only natural to expect that a great deal of exploration and study still needs to be done' [7].

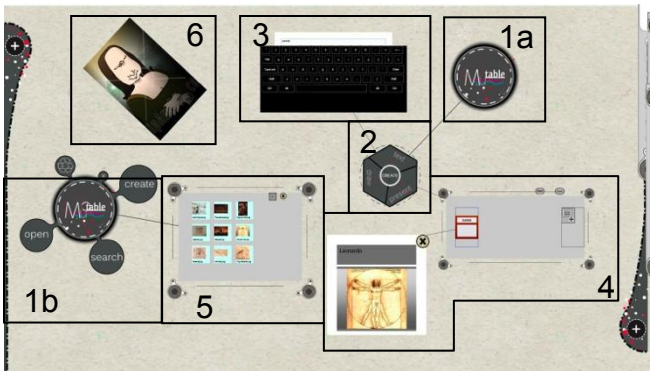


Figure 1. Main MuTable functionality. The functionality includes navigation elements and several widgets: 1. central ball menu (collapsed 1a. and expanded 1b.), 2. a submenu, 3. a typewriter tool, 4. a presentation creation tool (with a slide opened), 5. a file browser tool, 6. a piece of content (picture) opened.

Due to a particular mix of characteristics (e.g., a focus on productivity in a school context, see figure 1) and a lack of a clear gestural UI language, the design of the tabletop included a number of novel design solutions, sometimes inspired by other multi-touch developments, but not familiar to the general public. With an interface design based on much experimentation and little standards, an evaluation method capable of providing detailed feedback about non-standard user interfaces was needed. We chose for the Communicability Evaluation Method as a semiotic way into the analysis of multi-touch interfaces.

In the end, the application of CEM to the MuTable multi-touch platform was successful (it was published in a special issue on 'Developing, evaluating and deploying multi-touch systems' [2]). However, the cost at arriving at this outcome was unanticipatedly high. Although the method is not GUI-specific at first glance, a lot of time has been devoted to tailoring the method to the new multi-touch context. De Souza and Leitão [5] state that the Semiotic Engineering methods can be used in both 'technical and scientific contexts'. However, while the process of transfer proved to be worthwhile in the scientific perspective of the HCI research group I am working in, such an effort would probably have been considered a terrific waste of time and resources in an industrial setting, where the study focus would have been more on evaluation results. In other words, the time needed to do the study described in this paper was well-spent because its main outcome was the application of a research method to a new domain [12], rather than the specific evaluation results of the study.

THE TRANSFER FROM GUI TO NUI IN DETAIL

Developed before the boom in multi-touch application development, it is not realistic to expect the Semiotic Engineering Methods to be tailored to multi-touch situations. Moreover, Prates et al. [9] explicitly acknowledge that the 'method applies basically to single-user interfaces', leaving the possibility of introducing new tags and categories open for e.g. multi-user and artificial intelligence applications.

While we were expecting to add some extensions to CEM based on Prates et al. [9], the issues are more profound than making a few tweaks to the method. The changes were absolutely crucial for the application of the method to succeed. To be sure, the analytic power of the CEM framework depends heavily on the appropriateness of the low-level tags, and the higher-level interpretation taxonomy. We distinguish between two main difficulties: augmenting the method with useful, new tags and categories on the one hand, and dealing with the 'hidden WIMP bias' (windows, icons, menus, pointers) of the existing tags.

Creating New CEM Tags and Categories

Introducing new tags and categories for the use of CEM in new contexts is not always self-evident, despite the fact that Prates et al. [9] almost casually mention this possibility. In new contexts, it is difficult to anticipate beforehand which user problems 'are likely to occur', and need a new, separate tag, especially when applying the framework to new interface paradigms. This requires at least some experience with the technology and the test user group before one can make informed extensions to the CEM framework. Less appropriate tags and interpretation taxonomies will significantly reduce the framework's analytic power, by failing to point out the important 'communicative breakdowns' between designer and user.

Apart from the anticipation problem, it is a challenge to decide on what level to introduce new elements: as described above, CEM includes a low 'tag' level, and a higher 'category' level. In other words, after deciding to add an element, it is an equally important step to decide whether to introduce a series of new, 'dedicated' lower-level tags, or to re-use the existing tags, but add a new higher-level category to attribute them to. In situations where this decision is not clear-cut, the options need to be considered carefully, as they have important consequences for the resulting analysis.

We would like to offer the example of user problems related to gestural interaction. In specific, consider the

situation in which a user performs a gesture in such a way that it is not recognized by the system. The system does not recognize the gesture and remains inactive, and the user does not understand why his actions have no effect. One way of approaching this issue is to create a new tag for this type of gestural issue, as this interaction issues is a 'new' one, specifically relating to the use and availability of gestures. On the other hand, this gestural issue can be considered a slightly different version of an issue already present in the taxonomy. The original *Why doesn't it?* tag refers to a user struggling to make sense of an interface, because a specific part does not react as expected. In a GUI environment, this is typically the case when there is a breakdown in the user's understanding of the system, the user not understanding why the system is unable to complete the request. While the end result is the same in the gestural example (the user struggling to make sense *why the system doesn't*), the gestural issue is different: the request itself may be valid, but the user's execution of the request is not. CEM's original classification taxonomy does not allow for this type of issue, where the user's intentions are correct, but the user's actions are not recognized by the system.

For the gestural issue described above, we decided to go with the second option. Instead of adding a series of gesture-specific low-level categories, a higher-level category was added to the interpretation taxonomy. The existing low-level tags were interpreted more broadly to include gestural problems. The new *Gestures* category distinguishes between issues on a higher level: issues concerning communication about input methods are differentiated from those related to e.g. meaning assignment (interpretation of interface elements) and navigation. During analysis, this high-level category allowed to attribute a diverse set of breakdowns to the communication about gestures the designers had embedded in the user interface.

From the perspective of Woolrych et al.'s [17] description of methods as a set of resources, the adjustments to CEM (adding new tags and categories) were made based on local knowledge resources. Application-specific knowledge about multi-touch gestures was used to create the adjustments. However, the adjustments also needed to fit in the specific semiotic orientation of the method, focusing on designer-user communication. Therefore, the adjustments to CEM had to be considered carefully, balancing the local knowledge resources from the project with the theoretical specific orientation of the method.

WIMP Bias

Apart from the necessity to revise CEM for multi-touch purposes, the method's appropriateness was also decreased by another, more covert 'WIMP bias' of sorts. While none of the tags or categories explicitly refer to windows, icons, menus or pointers, applying the method to a new domain such as multi-touch showed that the tags implicitly target common GUI problems. I will offer a few examples of this bias.

One example of a WIMP bias can be found in the need to create the *Gestures* category described above. Semiotic engineering, developed roughly between 1993 and 2005, seems to take the interaction modality for granted. Mice and keyboards were the dominant input mechanisms at the time: these input mechanisms, and the way the input is transferred to, and visualized in the system is not taken up in the analysis. However, with a lack of gestural interaction standards in multi-touch interaction, the interaction modality is foregrounded, and analytic devices (tags or higher-level categories) targeting input mechanisms need to be added.

Another examples of the WIMP bias are the tags created for breakdowns due to multi-step navigation paths (implying deep menu structures, and embedded functionalities: the *I can't do it this way* tag), or on the modality of user interfaces (applications offering separate modes in which specific functionality can be available or not: the *Where am I?* tag). During a first inspection of the CEM tags and categories, the user-centered wording of the tags themselves (referring to the user in the first person) does not reveal a bias, and the method seems broadly applicable. It is only when applying it that the WIMP bias becomes apparent. In multi-touch multi-user systems, these WIMP-styled modality and menu structure-oriented tags are often not very relevant. Deep menu structures are generally avoided in multi-touch applications, and while every single screen in a multi-touch app can be seen as a separate mode [1], these modes are generally quite distinct, and offer only a limited set of functionality, reducing the risk of confusion. These tags were not removed from the taxonomy, to avoid overlooking these issues during analysis. However, in the analysis of the MuTable results, their occurrence was marginal.

In sum, specific elements in the method appeared to be tailored to GUI interactions, while the method needed extensions to cater for specific non-GUI interactions. These observations lead to the conclusion that CEM implicitly seems to be tailored to GUI evaluation. The creators of CEM do acknowledge that for specific applications, extensions to the method are needed [9]. As such, they explicitly present multi-user applications and artificial intelligence systems as outside the scope of the original CEM method. However, they do not mention interaction modality (keyboard and mouse vs. touch) when discussing the scope of the method. In this sense, CEM's scoping resources [17] are not entirely adequate.

Extra: What About Other Semiotic Engineering Methods?

After the above issues in applying CEM to multi-touch, we further explored the other main Semiotic Engineering method, the Semiotic Inspection Method (SIM), to see whether or not this method has similar issues associated

with it. SIM was not used to evaluate the MuTable interface, as SIM is targeted at *reconstructing* the designer's intended message from the interface. In this case, the technology was evaluated by the application's designers themselves. Therefore, as the 'designer's intent' was well-known to the evaluators, a SIM analysis would have lost at least some of its explanatory power.

A short, informal application of the method to the Windows 8 'Modern UI' (a 'touch first' interface) showed a similar WIMP bias as CEM. This bias can be found in the first three steps of the method: the analysis of 'metalinguistic signs', of 'static signs', and of 'dynamic signs'. These steps are aimed at '*deconstruct[ing]* the metacommunication message, allowing the researcher to inspect in great detail what and how the designer communicates with each type of sign' [5].

In typical GUI interfaces, a clear separation can be made between the three sign types described above. De Souza and Leitao offer the example of MS Word, in which metalinguistic signs (help texts, either within the application or online support material on Microsoft websites) can clearly be separated from static signs (menu items, toolbars,...) and dynamic signs ('what you see is what you get' WYSIWYG on-the-fly changes that happen in response to the users' actions). However, in contemporary (multi-touch) interfaces, this sign type division as a starting point for analysis becomes problematic:

- *Metalinguistic signs.* Increasing numbers of smartphone and tablet apps offer very little or no explicit user guidance, thereby all but eradicating the metalinguistic sign type.
- *Static signs.* In smartphone and tablet interfaces, 'static' interface elements are often few and far between. While mail apps typically do have a static menu bar in some screens, there are probably few apps that have static elements that recur across all screens. As stated above, this creates a situation in which every single screen in a multi-touch app can be seen as a separate mode [1]. The Windows 8 operating system, even without a user performing any action, constantly refreshes mail messages, pictures, etc. (see figure 2), reducing the static signs to a screen title, and fixed, empty boxes that are filled up with dynamic content.
- *Dynamic signs.* The SIM analysis of 'dynamic signs' is targeted at investigating *interaction* [5]. Dynamic signs are seen primarily as system feedback to the users' actions, as in the WYSIWYG example above. However, in contemporary interfaces, the system often refreshes and changes information without user action (see figure 2). This implies that the dynamic sign category needs further refinement into dynamic signs as a result of user interaction, and dynamic signs as a result of system-driven updates.

Figure 2. Windows 8 Modern UI.

The above analysis shows that SIM's analysis of contemporary touch interfaces can be expected to be very heavy on 'dynamic signs', with hardly any metalinguistic and static signs. While this observation in itself can be seen as merely a characteristic of touch interfaces, it becomes more problematic in the fourth step of the method (steps 1-3 being a separate analysis of all three sign types). Step four is based on the first three steps, and 'collate[s] and compare[s] the results of segmented metacommunication analysis. The aim is to detect inconsistencies and/or consistent relationships and patterns between elements collected in segmented analysis' [5]. While SIM treats the three sign types as equally important in the analysis, it is clear that this position is no longer tenable. The relative importance of the sign types needs to be reconsidered. Especially the dynamic sign type needs to figure more prominently in the analysis; the category also needs further refinement in subcategories to allow for an adequate analysis. Without such modifications, the semiotic inspection methods loses its much of its value due to a mismatch between the interface design under analysis, and the methodological tools.

CONCLUSION

The Semiotic Engineering case study presented in this paper shows that the transfer of a method from one context to another should not be taken lightly. In the Communicability Evaluation Method transfer, the changes to the method did more than 'augment' its applicability. For instance, the addition of a *Gestures* category to the method to cater for the specifics of multi-touch interaction was absolutely crucial for the research to succeed. As for the Semiotic Inspection Method, the short exploration presented above showed that a change in context to multi-touch applications also deeply influences the applicability of the method.

While the case study presented here describes a transfer of methods that are probably somewhat less familiar to the majority of HCI researchers, the results do indicate that the transfer of a method to a new context needs to be done with care. Even when no specific issues seem to pop up at first inspection, one needs to be watchful: hidden biases and other difficulties are not at all impossible.

REFERENCES

[1] Apple Inc. (2011). iOS Human Interface Guidelines. User Experience. Available from: /http://developer.apple.com/library/ios/documentation/userexperience/conceptual/mobilehig/MobileHIG.pdfS (accessed 20.07.11).

[2] Derboven, J., De Roeck, D., Verstraete, M. (2012). A Semiotic Analysis of Multi-Touch Interface Design: The MuTable Case Study. *International Journal of Human-computer Studies* 70 (10), 714-728.

[3] De Souza, C.S. (1993). The semiotic engineering of user interface languages. *International Journal of Man-Machine Studies* 39 (5), 753–773.

[4]De Souza, C.S. (2005). *The Semiotic Engineering of Human-Computer Interaction (Acting with Technology).* The MIT Press, London, UK.

[5] De Souza, C.S., Leitão, C.F. (2009). *Semiotic Engineering Methods for Scientific Research in HCI.* Morgan & Claypool, San Francisco, CA.

[6] Norman, D. (2004) Jacket blurb written for [4], quoted in: Norman, D., Design as Communication, http://www.jnd.org/dn.mss/design_as_communicat.html (accessed 08.03.13).

[7] Norman, D.A., Nielsen, J. (2010). Gestural interfaces: a step backward in usability. *Interactions* 17 (5), 46–49.

[8] Prates, R.O., Barbosa, S.D., de Souza, C.S. (2000a). A case study for evaluating interface design through communicability. In: *Proceedings of the 3rd Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques*, ACM, New York, NY. 308–316.

[9] Prates, R.O., de Souza, C.S., Barbosa, S.D. (2000b). Methods and tools: a method for evaluating the communicability of user interfaces. *Interactions* 7 (1), 31–38.

[10] Prates, R.O, Souza, C.S., Assis, P.S. (2001). Categorizing communicability evaluation breakdowns in groupware applications. *CHI-SA 2001*. Pretoria, RSA.

[11] Saffer, D. (2008). Designing Gestural Interfaces: Touch Screens andInteractive Devices. O'Reilly Media, Sebastopol, CA.

[12] Smith, S. P., Burd, E., Rick, J. (2012) Developing, evaluating and deploying multi-touch systems. *International Journal of Human-Computer Studies* 70 (10), 653-656.

[13] Vinh, K., 2011. Unnecessary Explanations. /subtraction.com Feb 01, 2011 (accessed 25.07.11).

[14] Westerman, 2008. How People Really Use the iPhone. Create with Context Research Report. Available from: /www.createwithcontext.com/how-people-really-use-the-iphone.html (accessed 15.01.12).

[15] Wigdor, D., Wixon, D. (2011). *Brave NUI World.* Morgan Kauffman, Burlington, MA.

[16] Wixon, D., 2008.The challenge of emotional innovation. UX Week. Available from: /http://vimeo.com/2893051S (accessed 19.07.11).

[17] Woolrych, A., Hornbæk, K., Frøkjær, E., Cockton, G. (2011): Ingredients and Meals Rather Than Recipes: A Proposal for Research That Does Not Treat Usability Evaluation Methods as Indivisible Wholes. *International Journal of Human-Computer Interaction* 27 (10), 940-970.

Adapting User Testing for Evaluating a Multiplayer Game

Marta Kristin Larusdottir
Reykjavik University
Menntavegur 1, 101 Reykjavik
marta@ru.is
Tel: +354-599 6200

ABSTRACT

In this paper we describe how user testing was conducted in two cases; first how the evaluation of a work related software was conducted and secondly how an evaluation of a multi-user game was conducted. The goal of the study is to analyse what adaptions needed to be made when transferring a rather typical approach of conducting user testing in the first case to conducting user tests on two prototypes of the multiplayer online game in the second case. For the analysis we describe, which, who, what, why, when, where and how the evaluation was conducted.

This study shows that it is vital to describe all these different factors involved when describing a particular case of user testing to better understand the particular case and the effect that the context of the evaluation had on how the results can be interpreted.

Our analysis of the adaptations needed of the user testing approach shows that all the factors analysed needed to be changed in some way. The fact that this was a multi-user software still being developed had the biggest influence on the adaptations.

THE TWO CASES OF USER TESTING

Two cases of user testing are described in this paper. User testing of a web based software used for scheduling work hours at a hospital and in a public authority organization is described in the first case. A rather traditional user testing using the think-aloud procedure with real users was conducted in that evaluation. On the contrary an adapted user testing approach was used in the second case where two prototypes of a multiplayer online role playing game were evaluated with user surrogates.

In the paper we first describe each of these two cases by analysing those according to the 6Ws and the H approach, like described for example in [1]. Furthermore we describe the major differences between these two cases and how these affected the user testing in the second case.

Case 1 – The Work Related Software

In the following the first case will be described according to the 6Ws and the H approach [1].

Which – Traditional User Testing Approach

User tests were conducted with 10 users on a new version of a web based software called Workhour, which is used to schedule work hours and look up monthly plans for shifts for example. The users were asked to think aloud during the task solving session of the user tests. The evaluations were conducted in the users own work environment. During the evaluation, the participants filled-in two questionnaires, and got tasks to solve in the software while being observed by two researchers.

Who – Real Users

There are four main user groups of Workhour; ordinary users that work on shifts and those that work regular hours. The other two main user groups are managers that work on shifts and those that don´t.

Five regular users took part in the tests, four working on shifts in hospitals and one working on regular hours in a state institution. There were also 5 managers that took part, two working on shifts at hospitals and three working regular hours, two at a state institution and one at a software company. The users are all familiar with older versions of *Workhour* and were chosen to be in the study as typical users of the system.

What – The Work Related Software Workhour

User tests were conducted on a new version of software called Workhour. This version had been running on a test database for two months when the think-aloud evaluations were conducted.

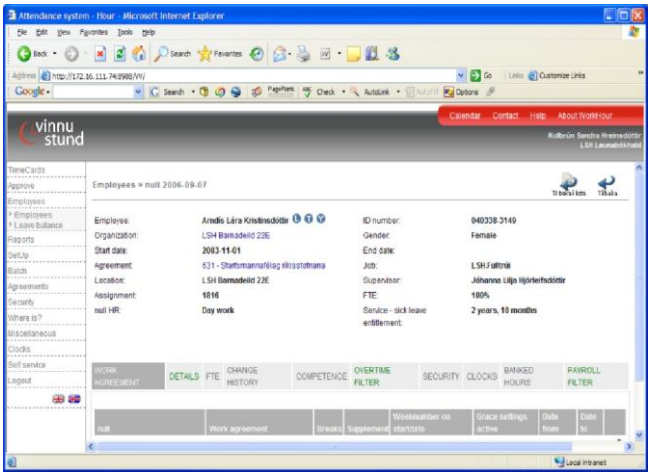


Figure 1: The evaluated version of the software

An old version had been in use for several years, but in the new version the user interface was changed extensively to being more menu driven interface but still rather traditional design, see figure 1.

The main user goal for ordinary users working on shifts for using the software is to check their monthly plan for shifts, ask for a day off and check if they have fulfilled all their work obligations for that month. The main tasks for regular users are asking for holidays and check if they have been too many hours off work. The Workhour system is very useful to managers, because they can do much of their organizing work in Workhour like check if all timestamps for their employees are correct, insert information about an employee that is sick and get an overview of how many have been sick over a particular period to name a few.

Why – For Researching Purposes - Summative

The reason for conducting the evaluation was to gather data for researching an approach for describing usability problems found in the user testing, described in [2] and to measure the user experience of the software, described in [3]. Furthermore, the results could be used for further design, but that was not the main purpose of the evaluation.

How – 6 Tasks and 3 Questionnaires

Ten usability tests were conducted by two usability specialists on the new version of Workhour. One of the usability specialists acted as a conductor and the other as a note taker and was responsible for the recordings of the sessions. In some sessions one of the developers of the system did attend as an observer.

Each user solved six or seven tasks in think aloud tests which were adjusted to their ordinary tasks. The total number of tasks in the study was 17. The tasks were made by one of the developers of the user interface that has good connections to the users. The users took part one at a time and were observed by two researchers in their own environment. In some of the tests a developer of the system was also observing the users.

The participants were asked to fill in the AttrakDiff 2 questionnaire [4] before and after the task solving session. First the participants were asked to answer the questionnaire according to their expectations to the new version of Workhour they would be trying in a minute. After each think aloud user test was finished the user filled in the questionnaire again and now the participant was asked to base his/her answers on the experience of using Workhour to solve the given tasks.

Where & When – In the Users Own Work Environment

The user tests were conducted at their ordinary working place, so a lot of contextual information was also gained. The user tests were conducted two weeks before the delivery date, so the software was almost finished at that time. Still there was time to fix serious usability problems.

Case 2 – Multiplayer Online Game

In the following the second case will be described according to the 6Ws and the H approach [1].

Which – The Adapted Approach

Two prototypes for a massively multiplayer online role playing game (MMORPGs) made in a prototyping tool called CADIA-Populus (see http://populus.cadia.ru.is/) were evaluated in case 2. The prototypes were made to compare two possibilities of moving avatars around in the game and two possibilities of starting a conversation between the gamers using the avatars. A within subject user test was conducted with 14 participants in total to test the user experience while using these two prototypes.

Who - The Participants

The users of the game will be people who like playing computer games that include natural behaviour of avatars.

We organized sessions of user testing with two groups of participants using the prototypes simultaneously, the first with 6 participants and the second with 8 participants. We needed to ask many participants to participate at the same time because this is a multi-user game.

The participants in the user testing were actually user surrogates, because the game had not been launched. I the first user testing session there were 6 participants, 3 males and 3 females whilst in the second session there were 6 males and 2 females. Participants were between 20 and 32 years old, mainly students or researchers in at the university.

What - The Game Prototypes

One of the principle strands of research at CADIA at Reykjavík University is the agent simulation of the human ability to communicate verbally and nonverbally based on detailed models of human behaviour and cognitive processing. As a part of that principle two prototypes called *System C* and *System S* were made in the prototype tool CADIAPopulus for evaluating dynamic group behaviour in virtual conversation.

To be able to evaluate different issues two prototypes were made *System S* and *System C.* The differences between those can be resumed in 6 points:

- *Input devices:* what input devices are used when controlling an avatar in the prototype.
- *Navigation:* in which way the user can let the avatar move around and explore the environment.
- *Attention visual cues:* how the system provides cues about the avatar's attention target.
- *Joining conversation:* in which way the user can join to a conversation.
- *Conversations:* how conversations are intended by the system.
- *Communication interface:* how the system presents communication features to users.

What - The System S Prototype

In *System S* the user controls his/her avatar forward and backward by the arrow keys. Additionally the user lets it rotate left and right using the respective arrow keys.

In order to start or join a conversation, the user had to approach another avatar, or group of avatars, and type something in the appropriated chat-box. All the avatars close enough to the user's avatar received the message. Therefore, conversations were intended as a bunch of people close enough to each other. In *System S* there were not attention visual cues to let the attention's target of an avatar stands out. Basically, the avatars in *System S* were agents with a very limited intelligence and not capable of being attentive or react to the environment, see figure 2.

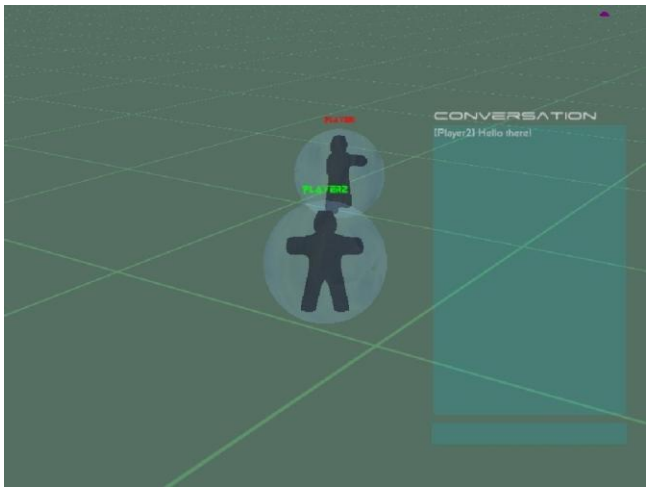


Figure 2: The prototype named System S

What - The System C prototype

In *System C* user controlled the avatar by means of a point-and-click interface using the mouse, see figure 3.

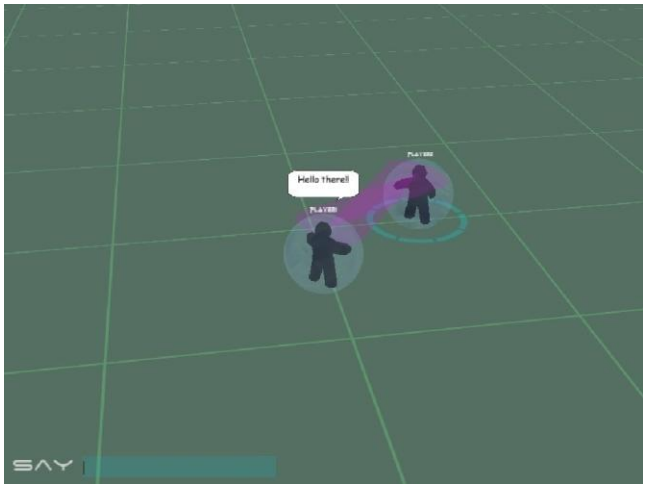


Figure 3: The prototype called System C

Touching the screen's border with the mouse pointer, the user could rotate the attention target and look around. In order to move the avatar, the user had to click on a desired destination point of the environment. Afterwards, the avatar would start to move in order to reach the chosen destination.

To start or join a conversation, the user clicked an avatar and typed something on the appropriate say-box. The message then appeared in a balloon above the avatar's head. If a chosen avatar was too far away, the avatar would start to move in order to get close enough to its target. Basically, an avatar needed to get attentive to another avatar in order to say something to it. Therefore it needed to move towards the target to let it fall inside its field of attention, which was a triangular region in front of the avatar. The field of attention pointed toward the avatar's attention target and changed colour and angular extension to visually describe different levels of attention focus. It is important to notice that an avatar was *aware of the existence* of another individual only if it did fall inside its field of attention and the avatar got attentive to it.

A Summary of the Differences Between the Prototypes

In *SystemC* avatars are smarter than in *SystemS* and are able to understand if they are in a conversation. Moreover, a chat-box will appear on the right side of the screen to let the user send and receive messages which belong exclusively to the conversation. The differences are summarised in table 1.

Table 1: Summary of the difference of the prototypes

	System S	**System C**
Interface Control	Arrow keys interface	Point-and-click interface
Navigation	Moving around the avatar controlling its trajectory	Setting a destination point
Attention visual cues	No attention visual cues	Field of attention with different angular extensions, orientations and colors
Joining Conversation	Approaching another individual	Clicking on another individual and typing something
Conversations	Based purely on distance	Based on a model of motivational social forces
Communication interface	A common chat-box	Context-based interface and balloon system

Why – For Redesign Purposes - Formative

The reason for conducting the evaluation was to gather data for redesigning the CADIA populus prototype and to measure how the dynamic behaviour in virtual conversions affected the users. Additionally we wanted to measure the user experience of using each prototype and compare the

differences. A redesigned version of the game was described in [5].

The goal was to evaluate the user experience of two issues:

1. *Avatar autonomous behaviour:* as an autonomous agent, each avatar is capable of reacting dynamically to other individuals. We used a model based on a potential field to drive the continuous action-reaction loop of an avatar participating in a conversation. We wanted to evaluate the validity of this approach in conceiving believability to users;

2. *Context-based interface*: our interface is able to understand in what social interaction the avatars are engaged and, consequently, to visualize parts of the interface without the user's intervention. We wanted to evaluate the comfort and the efficiency of this approach.

How – 6 Tasks and 3 Questionnaires

The participants filled in three questionnaires, solved three tasks while using each prototype and took part in a debriefing session. The participants were asked to use the prototype for communication only and not talk to each other orally. So actually they were asked to keep quiet while participating in the user testing.

In order to avoid biasing, the first group evaluated first *System C* and then *System S*; vice versa for the second group. In order to evaluate all the differences between the prototypes, we provided the users with three tasks to accomplish in each prototype.

At the beginning each user in a session were assigned a color and a team. The exact texts of the tasks for the first prototype were:

Task 1 - Find your way: You should log on to the prototype and find the mark on the floor that has the same color as you and stand on that mark.

Task 2 - Team up together: You are a member of the Angels team. There are 4 members in your team. Go around, start to talk with everybody and try to find out your teammates. When you find one of them, stay together.

Task 3 – Icebreaker – first prototype: Let us say you have the whole evening off next Friday night. Your team is supposed to spend the night together so you have to decide on what to do that night. Please be precise, so for example, if you decide to go to the movies you have to say which movie you want to see. (If you have any questions you can send one member to the oracle in the game and ask.)

The tasks for the second prototype had the same goals, but in task 3 the participants were supposed to schedule a weekend off together.

The participants were observed by two researchers where one of them was also a developer of the prototypes.

Before the evaluation each participant filled in a pre-evaluation questionnaire, collecting information on their background. After using each prototype the participants filled the AttakDiff 2.0 questionnaire [4] on the user experience while using the prototype and questions about the comfort and expressiveness of the prototype interface and the believability of the conversations. After using both prototypes, each participant filled in a comparison questionnaire in order to understand strengths and weaknesses of both prototypes. After that there was a debriefing session which was audio recorded. In addition to that all the conversations between the participants through the prototypes were logged.

Where & When – In Controlled Environment

The user tests were conducted in a problem class room at the university. In that room there were at least 30 desktop computers. The user tests were conducted on running prototypes, see figure 2 and 3. The results were used for redesigning purposes and the work continued on the systems for over a year after the evaluation.

THE ANALYSIS OF THE MAJOR DIFFERENCES

When analysing the adaptions that needed to be made to be able to transfer the rather traditional approach of conducting user testing of work related software to the evaluation a multiplayer online game a comparison was made on the factors described for each case. The comparison of the factors is summarised in table 2. In the following the similarities and major differences are described.

The similarities are mostly in how the user testing is conducted in that sense that there were: a) 6 tasks b) the Attrakdiff 2.0 questionnaire was used, c) background questionnaire was used and d) there was a debriefing session in both the cases.

The main reason for the differences is that in the second case we were evaluating a multi-user game, so we needed many participants to be there at the same time. This made the observation during the task solving session much harder to conduct. We tried to get an overview of what was happening by standing behind all the participants and look over the area where the participants were sitting, but it was impossible to see everything that each of the participants were doing. Therefore we needed to rely much more on the logging of what happened and their comments in the debriefing session than in the think-aloud user tests in the first case.

Additionally, the goal of the game was to use it for communication, therefore we wanted the participants to use the game as much as possible for that purpose and keep quiet during the evaluation. Also, if they had been using the game by themselves they would not be communicating orally, so we wanted to mimic the real context in this way.

This was a bit hard for the participants and especially one of the groups did not really keep quiet, so they started commenting to each other which is understandable, because there was not that much distance between the participants, so this was easy for them.

Table 2: Summary of the Differences of the Cases

	Case 1 Work Related Tool	**Case 2 a Multiplayer Game**
Which	Think- aloud user testing	Adapted user testing
Who	Real users from two user groups, ordinary users and managers.	Groups of user surrogates using the two prototypes simultaneously
What	a detailed prototype of work related software	2 game prototypes
How	6 tasks, background questionnaire and 2 versions of Attrakdiff 2.0, debriefing. Users observed one at a time by two researchers	6 tasks (3 for each prototype), background questionnaire, Attrakdiff 2.0, comparison questionnaire, debriefing. Users observed by two researchers
Why	For researching purposes mainly (summative)	For redesigning purposes mainly (formative)
When	A very detailed prototype 2 weeks before launching	2 less detailed prototypes in the middle of development
Where	In the users own work environment	In controlled environment

Furthermore, because we wanted to evaluate two prototypes in the second case, the evaluation procedure was different from the traditional one. The users solved three set of two similar tasks during the evaluation, but that did not seem to affect the evaluation much though. Additionally the users were asked compare the prototypes in the end, which was of course not done the traditional user testing case.

Finally, because the game was still being developed, we could not involve the real users in the evaluations and the users were maybe a bit homogeneous group and we could not evaluate the game in real settings because, we were evaluating prototypes that we needed to run ourselves.

CONCLUSION

In this paper we describe the adaptations that needed to be done on different factors while transferring a traditional approach of conducting user testing on work related software to conducting user testing on a multiplayer game. Modifications needed to be done in all the factors described to fit this new context of conducting user testing.

We can recommend using the 6W and the H for describing the context which the evaluation was conducted in. It resulted in a clear presentation of what was done and was a good tool for comparing the two cases in this paper.

ACKNOWLEDGMENTS

We would like to thank all the participants that participated in both these cases, described in this paper.

REFERENCES

1. S. Heim, The Resonant Interface - HCI foundations of interaction design, Pearson Education, Inc., 2008.
2. T. Thorgeirsson, and M.K. Larusdottir, Case Study: Are CUP Attributes Useful to Developers? , Proceedings for the COST-294 open Workshop, Downstream Utility: The Good, the Bad and the Utterly Useless Usability Feedback, Toulouse France, 2007, pp. 50 – 54.
3. J. Isleifsdottir, and M. Larusdottir: Measuring the user experience of a task oriented software. in: N.B. Effie L-C. Law, Georgios Christou, Mark Springett Marta Larusdottir, (Ed.), Proceedings of the International Workshop on Meaningful Measures: Valid Useful User Experience Measurement '08, Reykjavik Iceland, 2008, pp. 97-101.
4. M. Hassenzahl, AttrakDiff(tm). Internet Resource http://www.attrakdiff.de., 2007.
5. P. Claudio, H. Vilhjalmsson, and M. Larusdottir: Avatars in Conversation: The Importance of Simulating Territorial. in: A.e. al., (Ed.), Proceedings of the 10th International Conference on Intelligent Virtual Agents, Springer-Verlag, Berlin Heidelberg, Philadelphia, USA, 2010, pp. 336-342.

Mixed UX Methods Can Help to Achieve Triumphs

Leena Arhippainen, Minna Pakanen, Seamus Hickey
Intel and Nokia Joint Innovation Center, Center for Internet Excellence
P.O.Box 1001, 90014 University of Oulu, Finland
firstname.lastname@cie.fi

ABSTRACT
Each user experience test is a unique. Researchers need to find out how to introduce the research target (e.g. application, device, service) to the participants and what methods to use for capturing one's subjective experiences. Each test participant is a unique. Researchers need to know how to handle each subject, how to get them to express their experiences, wishes and needs verbally or non-verbally. Mixed methods test procedures can help researchers to 1) *introduce a test topic to participants step by step*, and 2) *learn about the users and their ways to express*, and 3) *catch user experience information piece by piece by utilizing different methods*. This paper presents practical examples of different case studies, where several mixed methods have been used together for getting a deeper understanding of users' experiences.

INTRODUCTION
User experience (UX) is an important factor for products' success [9, 28] and that is one reason why it has become a central target in product and service design [38]. During the last decade, the term user experience has spread everywhere in research and industry. Around the time of the millennium, the term user experience was like an ambiguous buzzword in product design and development [10, 12]. Before that, user experience was seen as a part of usability issues, but later it has been understood that even a product with good usability can cause negative experiences and vice versa [14]. A wide interest on UX during the last decade has changed the term from a buzzword to a considerable key asset of business and development. ISO 9241-110:2010 defines user experience as**:** *a person's perceptions and responses that results from the use and/or anticipated use of a product, system or service* [8].

Although the interest in UX in industry and academy is high, there is still a lack of methods on how to evaluate user experience [36]. Especially, there is a need to develop and use low-cost methods for UX evaluation and utilize the collected information in the early phase of the development process [37]. The aim of UX studies is to help in selecting the best design solutions, asses that the development is on the right track and if the final product meets the original UX targets [36]. In iterative and agile development processes, time- and cost-effective UX studies can give important UX feedback for the design [37]. During the last decade, UX research has matured and various conferences have been held around this topic for several years. In addition, there are several doctoral theses of UX [6, 13, 15, 16, 17, 18, 24, 25, 29], just to mention a few. In addition, there exist various UX professional communities around the world. One example is a UX community that maintain allaboutux.org site. One of their contribution is a white paper of UX which was a result from a Dagstuhl seminar, where 30 experts from academia and industry worked together to bring clarity to the concept of UX [31].

Although the interest in UX is high, there is still misunderstanding with the term and suitable methods. Based on the authors' subjective experiences, one typical misunderstanding relates to the question: *Can user experiences be studied without a functional application*? User experience should be evaluated *before, during and after the use* [36], and it can be evaluated with concepts in the early development phase [30]. User experience evaluation was one key topic in the CHI2012 conference, for instances, practical courses for UX evaluation were given by [32] and [7]. In the beginning of millennium, mainly traditional user research methods were used in UX research [21]. Since then, more user experience methods have been developed, for instance, AttrakDiff [11], iScale [19] and UX curve [20], just to mention a few. A collection of the current methods used in UX research are listed in the site allaboutux.org. Even thought UX research has matured, there are still missing knowledge and practical guidelines how to evaluate user experience, what methods to use and how to apply the gathered UX information in design. Especially there is not much knowledge of how to utilize creative mixed methods in different development phases.

This paper presents practical examples of the cases studies, where different UX methods have been used for getting a comprehensive understanding of user experiences. UX studies have been conducted in the research project with early concepts, existing commercial applications and functional prototypes.

USER EXPERIENCE STUDIES WITH MIXED METHODS
This section briefly introduces study cases and UX methods. We have studied three dimensional (3D) user interfaces from user experience point of view. In the research project we have conducted several UX studies with different examples using both approaches; early concepts and commercially available applications. Later we have conducted UX tests with our functional prototypes as well. In all studies, we have used the structured procedure with

predefined user tasks. However, in each study users have had a possibility to explore the test targets (applications, concepts) freely. In the following subsections, we used term *evaluation* when we have studied 3D UIs with concepts i.e. participants have *not used* any functional example. Accordingly by *experiment* term we refer to the test situation, where participants have *used* either existing application or our functional prototype.

A) TOY 3D Virtual Learning Environment Experiment

The UX studies of the 3D virtual learning environment were carried out with 30 pupils and students (10-18 years old) [3]. They used the 3D virtual learning environment from a PC or laptop with mouse and keyboard inputs. In this study we observed, for example, how users perceive and interact with 3D objects which are embedded into 3D virtual environment. The study procedure and methods:

- Classroom field test
- Recruitment: School, teachers
- Mixed methods:
  - Questionnaire before the experiment
  - Use of a functional prototype
  - Interviews + observations
  - Self-Expression Template: 3E-Method [35]

In this study, we found a lot of information of pupils' expectations and wishes towards 3D virtual learning environments. Interviews and observations elicited interaction experiences about the use situation. By the self-expression template, we found out how pupils saw them-selves as users of virtual learning environment. Creative methods can help to achieve Triumphs: users can express them-selves verbally and non-verbally.

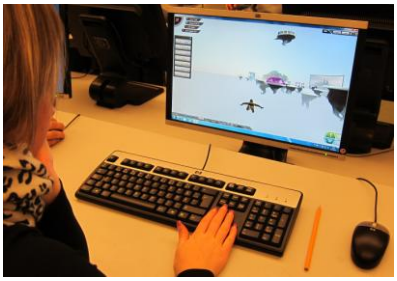


Figure 1. A pupil is using a 3D virtual learning environment in a classroom.

B) 3D Virtual Music Club Concept Evaluation

We conducted interviews with musicians and listeners, and based on that information we created scenarios and storyboards for a 3D virtual Music Club environment. Then we organized single interviews and focus groups, where participants had different tasks: storyboard walkthrough, comparison and selection tasks, and self-expression task. We had 23 participants, whose age varied from 24-56 years (Mean 34). Based on the results were modeled 3D virtual environments for a music context [2]. The Figure 2A presents one example page of the storyboard and B) the focus group session. The study procedure and methods:

- Single sessions / focus groups
- Recruitment: Personal invitations
- Mixed methods:
  - Background questionnaire
  - Interviews + observations
  - Storyboard walkthrough
  - Comparison and selection tasks
  - Self-Expression Template: Music Club [2]
  - Sketch ranking (best, 2nd and 3rd out of 10)

In this study, it was important first to introduce a topic to the participants, therefore we used early phase storyboards. When subjects were familiar with the topic, they were able to express opinions and ideas about 3D virtual music club. It was important to have selection and comparison tasks in order to get design directions. Because we used storyboards and sketches, participants perceived that they can influence on the future designs. This was rewarding for them. Based on the participant's drawings on the templates, we got 194 notes of ideas for the 3D virtual music club environment [2]. Based on this study, we made 3D models of the two different 3D music club environments [2]. Mixed method procedure in this study really helped us to find Triumphs: design directions.

A  B 

Figure 2. A) The 3D Virtual Music Club concept was evaluated in B) the focus groups using a storyboard walkthrough.

C) Hybrid 2D/3D User Interface Experiment

The use of the hybrid 2D/3D user interfaces on tablet devices was studied by conducting UX experiment for three mobile games and one demo map application [33]. In this study, we used various methods; observation, interviews, user tests, a customized version of the Product Reaction Cards [5] method and the Patio discussion forum [22]. We had 12 participants, whose age varied from 23 to 34 years. The Figure 3 presents one example of how a user interacted with the hybrid 2D/3D UI using touch gestures. The study procedure and methods:

- Single user test
- Recruitment: Patio online test user forum
- Mixed Methods:
  - Background questionnaire
  - Interviews + observations
  - Use of four existing applications [33]
  - Adjective selections (5/39)

In this study, we found all important issues by observing the use situation and interviewing during and after the use. Users' adjective selections supported the findings gathered

by interviews and new results were not elicited. However, the main benefit of using adjective selections in this case, was that they enabled to achieve quantitative information of users' experiences. Also it was easy and fast for users to select adjectives and then explain reasons behind the selections (experiences). Qualitative and quantitative methods together can help achieve Triumphs: a larger understanding of UX.

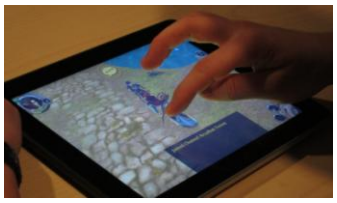
A  B 

Figure 3. A) A user is controlling his avatar and 3D game environment with two fingers (wrong gestures). B) A user is mimicking walk on the avatar (wrong gestures).

D) 3D Desktop User Interface Concept Evaluation

We created our first version of a 3D UI concept (Figure 4), and evaluated it with four users [1]. On the screen, a user can see a 3D space, where his/her phone applications and functions are located. The concept evaluation gave information about how users perceive 3D UI and what benefits such UIs could provide to them. The study procedure and methods:

- Single evaluation sessions
- Recruitment: Personal invitations
- Mixed Methods:
  - Background questionnaire
  - Non-functional tablet prototype
  - Concept walkthrough
  - Interviews + observation

This was very light and fast UX evaluation and only a few methods very used. However, this study elicited how we can get valuable UX information with low-fidelity prototypes for the future designs [1]. Triumph: cost-effective and time-saving.

Figure 4. A user is evaluating the 3D Desktop UI Concept using a non-functional virtual prototype.

E) 3D User Interface Concept Evaluation

In the early phase of our development process we conducted 3D UI concept design and evaluation phases [4, 26, 27]. In the evaluation, we presented ten different concepts to the participants. The concepts were shown as a non-functional virtual prototype on a tablet and on a laptop, or by paper-prototyping (Figure 5A). We also created a self-expression template for users to express their ideas about the 3D UI on a touch screen tablet (Figure 5B). We had 20 pair evaluation sessions with a total of 40 participants (age varied from 23 to 52, while the average was 35). The study procedure and methods:

- Pair sessions
- Recruitment: Patio forum, personal invitations
- Mixed Methods:
  - Background questionnaire
  - Interviews + observations
  - 2D/3D icon comparisons / tablet prototype [26]
  - Concept (1-4) evaluations/ tablet, paper, PC [27]
  - Concept (5-6) evaluations / tablet prototype
  - 3D depth evaluation/ Paper prototype [4]
  - Concept (7-10) evaluations/ PC, video, paper
  - Self-Expression Template: Paper Tablet

The procedure in this evaluation was very large and included several examples and tasks. The sessions lasted approximately 90-120 minutes. However, the topic was interesting for participants and the rhythm of procedure was balanced with the tasks, therefore participants were not exhausted after the session. Instead, they were surprised how fast the time had spent and how fun they had had. Because we studied 3D UIs from several perspectives, it was very important to use a mixed methods procedure with different types of concept examples (non-functional virtual tablet and PC prototypes, video and paper prototype). Likewise, it was critical that participants used the drawing template as a latest task. Thus, they were able to use all showed examples as a 'food' for drawing. Also this case elicited that a pair session is a good setup for experience elicitation and sharing, because the participant is expressing her/his experiences, wished and ideas to the other participant (etc. a friend, a mate), not only to the researcher.

A  B 

Figure 5. A) A moderator is introducing the concept (nro 7) to the subjects by paper-prototyping. B) Participants are expressing their ideas by drawing to the template.

F) S3D User Interface Experiment

In this study, we evaluated how users interact with the autostereoscopic 3D (S3D) user interfaces and what are their experiences [34]. In order to study user preferences for S3D UI, we selected an existing touch screen device with a glasses-free display, LG Optimus 3D phone and suitable 3D applications (S3D menu, S3D camera + gallery, Regina 3D Launcher and two 3D contact books). The evaluation followed a predefined structure, where users were asked to perform the detailed tasks relating to following activities:

A) Shoot a S3D photo and browse photos in a S3D gallery (Figure 6), B) Use a S3D menu, D) Use a Regina 3D Launcher and E) Use two different 3D contact books. After each activity (A-E), the subjects were asked to express their experiences verbally and by selecting 5 of the 52 adjective cards. The aim of using adjectives was to focus on the main topics of a S3D UI, and then discuss with users about their selections. Users were also interviewed and observed. We had four single and four pair tests sessions with a total of 12 users, whose age varied from 23 to 57 years, (Mean 30). The study procedure and methods:

- Single and pair user experiments
- Recruitment: Personal invitations
- Mixed Methods:
  - Background questionnaire
  - Interviews + observations
  - Use of four existing applications
  - Adjective card selections (5/52)
  - Comparison between 2D and 3D versions

In this study, it was really useful to use the adjective card selection method along with interviews and observations, because it again gave a quantitative data of the adjective selections, but also a deeper understanding of users' positive and negative experiences. For instance, users' positive experiences related to user experience issues, such as, *fun*, *entertaining*, *visually pleasant*, *exciting*, *innovative*, *new*, and *empowering*. Instead, negative experiences related mainly to usability issues, such as, *uncontrollable*, *unclear*, *useless* and *time-consuming*. In this case, mixed UX methods helped to achieve Triumphs: a larger understanding of the UX.

Figure 6. A) In the pair session, both users were asked to shoot S3D photos and B) then browse them in a S3D gallery and then select 5 out of 52 adjectives.

G) 3D Portal User Interface Experiment

In this study, we evaluated how users experience our 3D Portal UI and do they prefer it over to the 2D Tab UI solution. We captured user experiences by interviews, observation, adjective card selections, Likert Scale questionnaires and small selection tasks. In the study, we had 12 subjects, whose age varied from 20 to 40 years (Mean 28). The ratio between males and females was 1:1. The study procedure and methods:

- Single user test
- Recruitment: Personal invitations
- Mixed Methods:
  - Use of two demo applications
  - Background questionnaire
  - Interviews + observations
  - Adjective card selections (4/24)
  - Likert Scale questionnaires
  - Selection tasks

In this study we gathered information of users' preferences for 2D and 3D UIs by using three methods together and then comparing the findings. Based on the adjective card selections, Likert Scale answers and selection tasks we found out that the users prefer the 3D Portal UI over the 2D Tab UI, because of the richer user experiences it provides, such as visual pleasant, entertaining and fun. Both UIs were perceived to be equally easy and fast to use. In this case, mixed UX methods helped to achieve Triumphs: find out why 3D can be better than 2D.

Figure 8. A user is expressing her experiences by selecting 4 out of 24 adjectives after using our 3D Portal UI.

SUMMARY OF THE STUDY CHALLENGES

The aim of this paper was to introduce UX studies and show how we have investigated 3D user interfaces by utilizing mixed methods procedures with early concepts, existing applications and our functional prototypes.

The research topic has impacts on which kinds of concepts or applications can be used in the experiments. For example, if the topic related to professional activities (meetings, calendar event, contacts, etc.) it is not reasonable select examples from leisure contexts (e.g. games, Facebook). Also if the target device context is a touch screen tablet device, it is essential to take care that users are not thinking and comparing the research target to their PC use. The selected concepts and test applications or devices will influence on what methods can be used for capturing user experiences. For example, if a concept is very different from all what is familiar to a user, it is important to plan the evaluation sessions in a way that a topic is introduced to a participant little by little. Thus, the participant's understanding of the research target will increase and he/she can express experiences, wishes and ideas more precisely.

One main challenge in the UX research it to get participants to express them-selves in a way that researchers can gather, analyze and interpret the expressed UX information. By using mixed method procedure, researchers can enable subjects to express them-selves by different ways, for

instance, by verbal and non-verbal methods or by quantitative and qualitative questionnaires. In addition, the use of several methods together can help researchers to achieve a larger understanding of participants' experiences.

A CHECKLIST FOR UX STUDIES

Based on these studies, we have created a checklist, which can help to achieve Triumphs in UX research. Many of these bullets are relevant in all kinds of user studies, and a content of each task is dependent on how large the study is. However, in UX research, when a researcher is interested in user's subjective experiences, it is not good to control the task too much. Instead the researcher should let a user to 'lead' a conversation or use situation in order to get him/her to express experiences, wishes and ideas freely (and still take care that the test procedure is conducted as scheduled).

Planning

- **Plan/know what do you (want to) study**
- **Plan/know who do you (want to) study**
- **Plan a social context (Single, pair or group test)**
- Think how many participants do you want or need
- **Plan a physical + social context (where do you study)**
- Think what devices do you need
- **Plan what methods do you want to or can use**
- Think will you need assistants (camera, etc.)
- Take season and users' vacations into account
- Plan how to reward users
- **Plan how you will analyze the material**
- Plan a schedule for each test and a whole study

Preparation

- **Prepare materials (concepts, storyboards, etc.)**
- **Prepare / develop prototypes**
- Write questions and test them
- **Write a test procedure for you (moderator)**
- **Write a note form for you (scribe)**
- Write an introduction to participants
- **Make a background questionnaire**
- Make a permission form for using photos
- Check recording devices (video, audio)
- Start a recruitment of participants
- Acquire prizes/rewards
- Make guide signs for the test venue
- Conduct pilot test(s) and edit procedure if needed
- Reserve office material (pencils, tape, etc.)
- Book a test room / organize test venue
- Keep test users' phone numbers with you

Conducting

- Place guide signs
- Welcome participants
- Serve coffee, juice, water, etc.
- Remember to press REC
- **Take care of participants (support, do not force)**
- **Listen to the participant / ask additional questions**
- **Let a user to lead an experience discussion, even though you moderate**
- **Keep up a good feeling during the session**
- **Respect participants**
- Thank participants
- Reward

Like always, life can surprise, therefore be prepared or aware at least, that something can go wrong and a Tragedy is ready. Then, be creative, be positive, because each user is a worth of testing. Each UX study will give you new knowledge, a deeper understanding of user experience, and UX methods.

CONCLUSION

This paper presents practical examples of the cases studies, where different UX methods have been used for getting comprehensive understanding of user experiences. UX studies have been conducted in the research project with early concepts, existing commercial applications and research prototypes. Each time, the research target and test application and devices have influenced on the test procedure and methods. The whole process from a planning the test to the reporting results needs a lot of time, skills and lucky in order to achieve a Triumph. Therefore, we propose a checklist for UX studies.

REFERENCES

1. Arhippainen, L. Hickey, S., Pakanen, M. Towards a 3D User Interface in a Tablet Device Context. An iterative Design and Evaluation Process. In Proc. ACHI 2013, ThinkMind Press (2013), 47-52.
2. Arhippainen, L., Pakanen, M. & Hickey, S. Designing 3D Virtual Music Club Spaces by Utilizing Mixed UX Methods: From Sketches to Self-Expression Method. In Proc. MindTrek 2012, ACM Press (2012), 178-184.
3. Arhippainen, L., Pakanen, M., Hickey, S. and Mattila, P. User Experiences of 3D Virtual Learning Environment. In Proc. MindTrek, ACM Press (2011). 222-227.
4. Arhippainen, L., Pakanen, M. and Hickey, S. Studying User Experiences on Depth in 3D User Interface by a Paper Prototype as a Part of the Mixed Methods Evaluation Procedure. In Proc. ACHI 2013, ThinkMind Press (2013), 35-40.
5. Barnum, C. and Palmer, L. More than a feeling: understanding the desirability factor in user experience. In Proc. CHI2010, ACM Press (2010), 4703 – 4715.
6. Battarbee, K. Co-Experience. Understanding user experiences in social interaction. Dissertation. 2nd ed. University of Art and Design, Helsinki. 2006.
7. Bernhaupt, R. User Experience Evaluation in Entertainment and Games. CHI Course. 2012.
8. ISO DIS 9241-210:2010. Ergonomics of human system interaction - Part 210: Human-centred design for interactive systems. International Standardization Organization (ISO). Switzerland.
9. Jordan, P. W. Designing Pleasurable Products, An introduction to the new human factors. New York, Taylor & Francis, 2000.
10. Forlizzi J. and. Ford S. The Building Blocks of Experience: An Early Framework for Interaction Designers. In Proc DIS, ACM Press (2000), 419–423.

11. Hassenzahl, M., Burmester, M., and Koller, F. 2003. AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. In J.Ziegler & G. Szwillus (Eds.), Mensch & Computer 2003. Interaktion in Bewegung (187-196). Stuttgart, Leipzig: B.G. Teubner.
12. Hassenzahl, M. and Tractinsky, N. User Experience – a research agenda. Behaviour and Information Technology (2006) 25(2): 91–97.
13. Häkkilä, J. Usability with context-aware mobile applications. Case studies and design guidelines. Doctoral thesis. University of Oulu. 2006.
14. Jokela, T. When Good Things Happen to Bad Products: Where are the Benefits of Usability in the Consumer Appliance Market? ACM Interactions, 2004, XI 6: 28-35.
15. Jumisko-Pyykkö, S. User-Centered Quality of Experience and Its Evaluation Methods for Mobile Television. Doctoral thesis. Tampere University of Technology. 2011.
16. Kaasinen, E. (2003) User needs for location-aware mobile services. Personal and Ubiquitous Computing 7(1): 70–79.
17. Kaikkonen, A. Internet on Mobiles: Evolution of Usability and User Experience. Doctoral thesis. Helsinki University of Technology. 2009.
18. Kankainen, A. Thinking Model and Tools for Understanding User Experience Related to Information Appliance Product Concepts. Doctoral thesis. University of Technology, Helsinki. 2002.
19. Karapanos, E., Martens, J.-B. and Hassenzahl, M. On the Retrospective Assessment of Users' Experiences Over Time: Memory or Actuality? CHI'10 EA. ACM Press (2010).
20. Kujala, S., Roto, V., Vaananen-Vainio-Mattila, K., Karapanos, E., Sinnela, A.: UX Curve: A Method for Evaluating Long-Term User Experience. Interacting with Computers (2011).
21. Kuniavsky, M. Observing the User Experience – A Practitioner's Guide to User Research. Morgan Kaufman, Elsevier. 2003.
22. Laizane, S. and Haukipuro, L. Preliminary experiences with the online forum PATIO in a multi-contextual Living Lab environment. In Proc. ISM 2012 workshop Innovation Through Social Media. 2012, 24-31.
23. Law, E., Roto, V., Hassenzahl, M., Vermeeren, A. and Kort, J.: Understanding, Scoping and Defining User eXperience: A Survey Approach. In Proc. CHI 2009, ACM Press (2009).
24. Luojus, S. From a momentary experience to a lasting one. The concept of and research on expanded user experience of mobile devices. Doctoral thesis. University of Oulu. 2011.
25. Mattelmäki, T. Design probes. Doctoral thesis. Univ of Art and Design, Helsinki. 2006.
26. Pakanen, M., Arhippainen, L. & Hickey, S. Design and Evaluation of Icons for 3D GUI on Tablets. In Proc. Academic MindTrek 2012, ACM Press (2012), 203-206.
27. Pakanen, M., Arhippainen, L. and Hickey, S. Studying Four 3D GUI Metaphors in Virtual Environment in Tablet Context. In Proc. ACHI 2013, ThinkMind Press (2013), 41-46.
28. Pine, B.J. and Gilmore, J.H. Welcome to the Experience Economy. Harvard Business Review. July-August 1998.
29. Roto, V. Web Browsing on Mobile Phones – Characteristics of User Experience. Doctoral thesis. University of Technology, Helsinki. 2006.
30. Roto, V., Rantavuo, H., Väänänen-Vainio-Mattila, K.: Evaluating User Experience of Early Product Concepts. In Proc. DPPI'09. Compiegne, France. (2009)
31. Roto, V., Law, E., Vermeeren, A., and Hoonhout, J. (eds). 2011. User Experience White Paper – Bringing clarity to the concept of user experience. Outcome of the Dagstuhl Seminar on Demarcating User Experience, Germany. http://www.allaboutux.org/uxwhitepaper
32. Roto, V., Vermeeren, A., Väänänen-Vainio-Mattila, K., Law, E. L-C., Obrist, M. User Experience Evaluation Methods: Which Method to Choose? CHI Course 2012.
33. Salo, K., Arhippainen, L. and Hickey, S. Design Guidelines for Hybrid 2D/3D User Interface on Tablet Devices. A User Experience Evaluation. In Proc. ACHI 2012, ThinkMind Press (2012), 180-185.
34. Sunnari, M., Arhippainen, L. Pakanen, M. and Hickey, S. Studying User Experiences of Autostereoscopic 3D Menu on Touch Screen Mobile Device. In Proc. OZCHI 2012, ACM Press (2012), 558-561.
35. Tähti, M. and Arhippainen, L. 2004. A Proposal of collecting Emotions and Experiences. Proc. HCI 2004, volume 2. 6-10 September, 2004. Leeds, England.
36. Vermeeren, A., Law, E., Roto, V., Obrist, M., Hoonhout, J. and Väänänen-Vainio-Mattila, K. User Experience Evaluation Methods: Current State and Development Needs. In Proc. NordiCHI 2010. ACM Press (2010), 521-530.
37. Väänänen-Vainio-Mattila, K. and Wäljas, M. Developing an expert evaluation method for user eXperience of cross-platform web services. In Proc. MindTrek 2009. ACM Press (2009), 162-169.
38. Väänänen-Vainio-Mattila, K., Wäljas, M. Evaluating User Experience of Cross-Platform Web Services with a Heuristic Evaluation Method. Int. Journal of Art and Technology (IJART), 2010.

Developer–User Social Distance as a Guiding Concept for User Involvement

Mikael Johnson
Helsinki Institute for Information Technology HIIT,
Aalto University
PO Box 15600, 00076 Aalto, Finland
mikael.johnson@aalto.fi

ABSTRACT
Social media changes the conditions for user involvement in service development. Active user communities, fast paced iterative development after market launch, developer access to users' digital trails, and low cost software distribution are well known facets that bring substantial changes. This paper articulates how these and other changes shape user involvement routines, including usability evaluation and user experience design and evaluation methods, based on an in-depth case study of an over decade-long service development in industry, Habbo Hotel by Sulake Corporation. As a benefit of its longitudinal approach, this study brought a neglected slowly changing contextual aspect in focus: *developer–user social distance*. The argument is that developer–user social distance could become a guiding concept for user involvement, thus supporting the transfer and adoption of methods between design contexts.

Author Keywords
User involvement, methods, routines, context, social media

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

INTRODUCTION

Social media is all but a clear-cut case for interaction designers. Despite many interesting phenomena—e.g., user-created content, user-driven or participatory web, wikinomics [1, 7, 12, 15]—we do not know whether it is feasible to design social media with a user-centred design process, nor to which degrees social media is 'user-driven' or 'participatory'. Contributions to the social media user experience provide some details [2], but little design process guidance in relation to when which usability and user experience design and evaluation method could be applied. Blind spots in the literature include assuming the design context to be a 'one-off project', the unclear role of developers' informal engagement and personal experience, and user involvement after market launch. [5, 13, 14, 16]

As HCI professionals face the challenge of the expanding scope of interaction design, a pertinent question is how lessons can be transferred between cases. What are the relevant aspects of generalization, and do the traditional frames of usability evaluation methods hold when the research field expands? Vice versa, can traditional wisdom guide social media development and bridge differences between work and leisure?

In this paper, we explore a large-scale social media case. The empirical data was collected during 2003–2010 by the author, who studied the vendor organisation in several research projects. This online service started out with the developers developing for themselves and their friends, and, like other success stories, found that there was a more general demand for their service. Over time, significant changes in the relationship between developers and users occurred, and the forms of user involvement transformed accordingly. The lessons learned from this case are distilled with reference to HCI method adoption themes: understanding which, why, who, what, when, where and how users were involved in design. The contribution of this paper is a new guiding principle for user involvement, developer–user social distance, to support HCI method deployment.

HOW SOCIAL MEDIA CHANGES USER INVOLVEMENT

User involvement is typically framed in two ways: either how to manage a particular interaction situation with one or more users, or the planning process when one decides how to approach users. We know a lot about particular methods to learn about users—interviews, observation, surveys, focus groups, field visits, cultural probes, and so on—and a fair deal about which factors drive the use of a method in research settings. However, we know very little about the factors that drive the selection of methods use in the long run, in a series of projects in product or service development organisations.

Users are often assumed to contribute through the means of user research, user requirements definition, context of use models, use case and scenario modelling, persona descriptions, and evaluation with users. However, anecdotal evidence from social media startups suggests that many developers did not start with typical user-centred design methods, but rather by developing the service for themselves [4]. On the other hand, many prominent social media companies have hired user experience designers and user researchers to learn from the users of their services. These weak signals intrigue us and lead to the question, what exactly is the role of users and user involvement methods in the design of social media?

Social media is here treated as a computerisation movement, a concept by Kling and Iacono that considers three components that interact with and shape each other: technological frames, public discourse, and organisational practice and use. First, specialised and mainframe computers, then mini and micro computers, computer networks, and related software were taken into use by organisations for different reasons (productivity, democratisation, collaboration). This time it is a combination of useful and usable computer-based technologies for consumers, services for groups of people, business model innovations, and active content-sharing users that is changing society. [2, 11]

Social media is relevant to design contexts in at least two different ways. First, all product and service developers can benefit from various collaborations with users through social media. Second, development of social media services for consumers can be delineated in terms of (1) software business (e.g., low cost of construction, modification, distribution, considerable development after market launch, unconventional revenue models), (2) features for group communication that make social media more than groupware (e.g., open-ended messages and other user-created content, support for a collection of groups, and high degree of awareness of other users' activities), and (3) use and users: active user communities, peer production, and high degree of voluntary use. [11] The case reported here exemplifies this latter setting.

In usability and user experience design and evaluation methods so far, a number of factors that influence method use have been proposed. The ISO standard entitled Usability methods supporting human-centred design [6] lists a number of factors in an appendix that influence method choice. In the standard these factors are structured by software lifecycle, project, user, task, and product characteristics, as well as, available skills. For instance, does the designer have access to users or are they too remote—geographically or organisationally? What ergonomics/human factors skills does the design team have? And, how much time and money is available?

While some of these factors are related to the context of use and some to the development context, not all aspects have been fully developed. For instance, the design team's expertise is only visible through human factors skills, while developers' familiarity of the use context is not noted. In this paper we also engage with two other concerns: the role of informal engagement between designers and users (in contrast to formal methods), and new sources of data about users in social media contexts. These concerns have been raised in recent debates on the design context and designer subjectivity. [13, 14]

In the following we will be looking at the relation between design and use context, and, based on the observations, propose a guiding principle for user involvement.

CASE HABBO: DATA, METHODS AND OVERVIEW

Habbo is one of the oldest and most popular social media services where children and teenagers meet, socialise, and play many types of games. The service is designed as a virtual hotel that encourages players to get a virtual hotel room, purchase virtual furniture and decorate their hotel room to their own taste and as a meeting place for games and socialising with other players (Figure 1). During 2003–2010 the service expanded from 4 localised hotels and 1 million monthly users to 11 language versions with 15 million monthly users from over 150 countries. Instead of an entrance or a monthly fee, the business model is free-to-play—revenue is based on micropayments and advertising in the hotel. Players, called 'Habbos', are encouraged to create their own objectives alongside chatting, room decoration, and meeting friends. Most of the teenage players log on after school, and according to Sulake, the developer company, on average they spend around forty-five minutes per day in the hotel or on its related discussion forums.

Our data was gathered both from developers and users during 2003–2010 through a multi-method approach with varying intensity during eight years and has been reported in detail in a PhD thesis [11]. The research started in the fall of 2003 with pilot interviews and participant observation in Habbo user communities. During 2004 the focus was on visitor profiles, studied through a survey that reached 10 000 users, and online texts written by Habbo users on websites, blogs and in discussion forums—so called Habbo

Figure 1. Idyllic Image From Ad for Habbo by Sulake in 2006.

fansites—to understand the consumption and user activities in Habbo. In 2005 ten theme interviews with Habbo developers and three focus group interviews with altogether twelve Habbo users were organised. In 2006 the author participated as researcher in the development of customer feedback methods at Sulake. From 2007 the research has concentrated on analysis, trying out new features in Habbo and keeping up-to-date through additional interviews with Sulake developers.

The data analysis proceeded in multiple waves over the years. A survey provided quantitative information of the use of Habbo. Analysis of texts written by Habbo users on fansites explored different Habbo consumption styles, popular activities, and hotel history. The topics of the user interviews were their participation histories, changing motivations, and meanings given to membership and

reference groups in Habbo. Taken together, these bodies of data provide us with an excellent view of the varying forms of interchange and dialogue between the users and developers of this social media service. This case is representative beyond its target group and games to social media in general, because of similarities in software business, group communication functionality, and active user communities. [8, 9, 10, 11]

The following account of deployed methods extends beyond standard usability and user experience evaluation methods to include other encounters between developers and users that serve similar functions in providing developers with information about the use and users.

KEY THEMES IN HABBO HOTEL SERVICE EVOLUTION

What Sulake–Habbo consists of has changed significantly over the years. Habbo started as a pet project for a few developers and their friends, grew to become a popular online world among new media people and within a few years it became mainstream for a teenage target group. Technical, economical, and organisational bottlenecks were solved so that the service could grow and scale up to become a transnational service. We group the service evolution into five stages (Table 1).

Stage	Years	Monthly users	Hotels
Concept	1999–2000	< 10 000	1
Beta	2001–2003	< 1 million	4
Expansion	2004–2005	1–5 millions	16
Complexity	2006–2007	5–10 millions	19
Competition	2008–2010	10–15 millions	12–18

Table 1. Habbo Service Evolution

Concept refers to the first prototypes in 1999 and 2000: Mobiles Disco, Lumisota, and Hotelli Kultakala. At this time, the development resources were minimal as the two founding developers created the first prototype on their free time after work and during weekends. Beta refers to the time period between 2001 and 2003, when much of the basic functionality was completed. Internationalisation started through a UK partnership, followed by a Swiss partnership. Expansion refers to 2004–2005 when the product was packaged so that it made a roll out possible in more than 10 new countries during one year. Before that different code was used in different countries. Complexity refers to 2006 and onwards when the product was extended to a social networking service. Competition reflects the increased amount of social media services for children and an increased teenage adoption of Facebook.

Strategy Change Due to Shifts in Developer–User Social Distance

During this service evolution, what was designed and developed changed. The concept stage started with making Habbo a cool hangout online and the developers were developing the service for themselves, their friends and their new media colleagues. Developers had easy access to users in the Finnish user community as developers could log on to Habbo and check what was going on. While the developers also used the service themselves, the informal engagement with the user community gave the developers a good implicit understanding of the users. Various informal evaluation practices, such as the slogans ‘easy access, easy play’ and ‘where else’ which had a shared meaning among the developers, guided the design early on. During the first year developers received abundant e-mail feedback by users, which became a handy source for design inspiration for the developers, who used to return to it periodically to browse for good ideas.

During the beta stage, designers focused on typical usages and the changing target group. With too many users to keep track of, the developers turned to typical usages: logging in, learning to navigate in Habbo, connecting with others, creating a room etc. The user base extended to a younger demographic and an age gap emerged, which had fundamental consequences to the service. Means for safe playing were implemented and the parent of the user became a key stakeholder in website communication. The fading insider perspective necessitated market and user studies to understand the new target group and a typology to communicate it. A back-end service that kept track of furniture sales across hotels was developed, allowing a comparison of Habbo features on the basis of their economic performance, not only based on functional or aesthetic properties. Fansite discussion forums were an additional important source for design inspiration. These means to learn about users were used to compare user bases in different hotel countries.

As the monthly number of users approached 1 million in four different countries, hotel-specific country organisations emerged as intermediaries between end-users, volunteers, and the increasingly centralised game development. These country offices would take care of the local technical configuration of the hotel, community management, customer support, local campaigns, and advertising.

Active Users and Emergent Developer Strategies

A key factor for service success in the early stages of its lifecycle was the emergence and continuous management of the fansites and volunteers programme. Already from the start of the service, groups of active Habbo users teamed up and created Habbo-themed websites in the form of blogs, online magazines, or discussion forums. These fansites emerged around all Habbo Hotels in their respective countries or language regions. They varied in size and temporality, from small sites with a few web pages that operated for a few weeks to the biggest fansites with hundreds of thousands of page views, readers in more than one country and that operated for many years. While most fansites remained fairly underground phenomena, the more popular ones got recognized by Sulake as "Official Habbo Fansites". This programme of giving special status in the

community to certain fansites started after the first three years of the service, during which the developers had operated their own official online fanzine, which also served as a model for later user-produced fansites.

The Habbo fansites served important community-building purposes, as they were run by active users and subgroups formed around them. For instance, they complemented the official website, strengthened the governance policies of the producer, reproduced and reinforced social positions (like potential Habbo career paths or legitimized user groups), and improved the Habbo users' awareness of the fan cultures around Habbo. In this case the developers could benefit from the massive amounts of online discussion about Habbo, which transformed qualitative inquiry in user research from being a prime means to gather data to being a means for source critique of what the users write about Habbo, and taking actions to fill in the gaps and skews in the users' online reports.

For the first five years of Habbo, Sulake leaned on volunteers to moderate the online activities. Volunteers were called 'Hobba' and their function was to mediate in conflicts, send warnings to misbehaving users, kick them out of the hotel rooms, or ban them from the hotel. To share experiences and moderation policies, the volunteers created an online forum for themselves. Along with the internationalisation and more organised volunteer management, Sulake started hosting a local volunteer forum per hotel country. The volunteers soon got an important role as mediators of user opinions: the developers knew that as the volunteers spent the most time in the hotel, they were always the first to know about the current user concerns, wishes and emergent activities.

During the expansion stage, many development practices became more formal and cost-efficiency became more important. As the organic beta testing phase changed into a more controlled release management process, Sulake started piloting the release for one month in one hotel country, before diffusing the release to other hotel countries. Playability testing was used to assess various playability aspects, such as gameplay, game mechanics, appearance, sound, and social playability. During 2004–2005, focus groups were conducted to evaluate the applicability of Habbo pixel style graphics and use of colours for the Asian market. The target group of the first usability evaluation was new users and business critical service features.

As some hotel communities grew larger, pressure emerged for customer service to automate their responses. For instance, in a country with several hundreds of thousands of users, a new feature might spawn several thousands of inquiries per day. In 2005, a new customer relationship management system was introduced. It featured a set of standard questions and responses, which reportedly reduced inquiries by 90 percent. In 2005–2006 Sulake brought the moderating function in-house, by employing moderators in their country offices. The volunteer program changed, and experienced Habbo users could apply to become so called Habbo eXperts, who did not have moderating powers anymore, but could get into a room that was full.

Cumulative and Strategic User Categorisations

In the complexity phase, many ways of understanding the diversity of the users were developed. In an effort to gather systematic feedback before the implementation of new features, Sulake recruited 200 volunteers in one country to form an online panel. Market research surveyed users' lifestyles, favourite brands and media usage patterns across different countries. User and group homepages and dynamic indexing systems (e.g. tags) served both the communication between users and the developers' interest in learning about the users. In 2008 the volunteer program changed again, and eXperts became Habbo Guides, who volunteer to welcome new users and explain Habbo's features. In 2009, Guide "Bots" were introduced, answering basic questions about Habbo.

In the competition stage, global competition and multi-sided business grew in importance. In 2009, when the Personas method was implemented in a data-driven fashion, the focus became to ensure that six persona descriptions should reflect the growing and declining market areas as well as have an even gender and age spread. The idea was that developers have an updated reference to the goals and needs of Habbo users at hand, which could inform design solutions and evaluations. The process of learning from surveys had been significantly developed with the aid of automation and web analytics techniques.

Table 2 summarises the above observed user involvement routines, and pinpoints their first occurrence to service evolution stages.

Stage	**User Involvement Routines, First Occurrence**
Concept 1999–2000	Avatar activities, Developers as users, Informal evaluations, E-mail feedback, Volunteers
Beta 2001–2003	Volunteer forum, Weekly newsletters and polls, Fansites, Official Fanzine, Summer meetings, Sales statistics, Customer service
Expansion 2004–2005	Market survey, Focus groups, Usability evaluation, Playability testing, CRM system, Release pilots
Complexity 2006–2007	Online user panel, Global youth survey, User and group homepages, tags
Competition 2008–2010	Data mining, Automated surveys, User experience testing, Personas

Table 2. First occurrence of User Involvement Routines by Service Evolution Stage [8]

SLOWLY CHANGING CONTEXTUAL DETAIL

When studying the accumulation of user knowledge in the development organisation over several years, it became apparent that project phases did not structure the deployment of usability and user experience methods. Instead the question became one of turning attention to shifts in developer–user social distance and a number of other contextual factors.

Developer–User Social Distance

The concept of developer–user social distance emerged from this case, as I had the opportunity to learn about a social media company's user involvement practices over several years and I became witness to a gradual, but significant, change in how users were involved in design. With an increasing number of users, more features, and geographic expansion of the service, also the diversity of use practices increased. The younger demographic of the users brought increasing differences between developers and users. Developers' active participation in use communities decreased, and volunteer users' participation in development and moderation waned. The role of the fansites changed as certain discussions about Habbo could be carried out in the developer-provided forums.

I conceptualised the above changes in the development as changes in social distance between developers and users. As developed in [11], shifts in developer–user social distance refers to changes in uncertainty and familiarity of the other group's practices, resulting from a combination of changes in 1) diversity of use practices, 2) differences between developers and users, 3) direct developer participation in use practices and vice versa (direct user participation in development practices), and 4) indirect contact between developers and users through both social and technical mediators.

Developer subjectivity, for instance, a developer's own use of a particular product or service and resulting first-hand experience, is poorly considered in guidelines and other advice on user involvement. Much writing on user involvement starts with the assumption that a developer is not a representative user and can therefore not trust his or her own gut feelings with respect to design choices. The other extreme opinion is also common, that is, developers are competent members of a community of practice and their personal experience is perfectly representative. [11, 13] In contrast, I argue that developers can lean on their ideas about use and experience of use, but that it depends on how familiar the developers are with the users and the use practices—what I call here the developer–user social distance.

This case is an example of self-centred design being adequate, but within certain limits. To convey the limits, the sensitising concept of developer–user social distance is proposed. As long as the distance is small, one can posit that self-centred design and informal user engagement can work, but as soon as the developer–user social distance grows, more effort is needed in terms of user involvement to bridge the emerging gaps. It also works the other way around. In many product and service design cases, the initial developer–user social distance is broad; however, as users engage in development and personal contacts develop, the developer–user social distance decreases, which then opens up possibilities for the use of more informal, potentially lighter and more first-hand, methods.

Other Aspects of Importance

Developer–user social distance is intended as a guiding concept, or a shorthand abstraction, for designers to communicate many complex relations between design and use. However, also a number of other contextual factors shaped the deployment of usability and user experience methods, too many to be covered in detail here. The following gives a brief overview to provide empirical support of their significance with respect to method deployment (cf. [16]):

Organizational specialization: When organisational specialisation increases, which tends to happen when organisations grow, more effort is needed on communicating knowledge about users and their use practices within the organisation, as not all managers and developers can have deep knowledge about users and use practices. This makes reports from user studies, use cases, scenarios, and other user representations more relevant.

Degree of business/mission criticalness: Login, registration, payment processes, and other factors enabling a low threshold of use are critical parts of most services, which need to work optimally. Sulake focused its first formal usability evaluation on these processes. On the other hand, less important features can stand more bugs or longer fixing times. Sulake left low-priority features hanging for a while.

Project scope & openness of design space: User feedback and use practices have most influence on the features that are under active development. Early on, emergent use practices and user feedback were significant, for instance, in the development of furniture ownership rights and their sharing, navigation between rooms, furniture trading mechanisms, moderation, and online discussion about Habbo. In 2006–2007 the service concept was broadened with social networking features and user feedback could influence those developments.

Feature-specific use variance: Assessing relevant variance in use practices is significant for fitting a technical feature to social practices. For instance, login, registration, and particular payment options are features with use practices that are tightly scripted with little degree of freedom. On the other hand, decorating a room and moving about in Habbo are very open-ended use practices. For open-ended use practices, technical flexibility is key and user research methods that can tackle open-endedness (observation, interviews, data mining server logs with machine learning algorithms). For tightly scripted use practices, clear interaction sequences are key, as are user research methods

with a high degree of control and a priori definitions, like A/B testing and quasi experiments, for instance.

User-Generated Content and User-Owned Services. Social media settings accentuate the organising of user communities and peer production after market launch. Active user volunteers can fill in where a service developer company has no resources. Key questions to the organization of user possibilities in influencing service evolution include who hosts, maintains and controls the rights to activities and outcomes of user-run, developer-run, or interconnected third-party blog/forum resources and services.

Digital trails. In social media contexts, developers have easy access to online user action, so whenever a question of uncertainty comes to mind, a developer can just log on and check what users are doing and writing about just that topic. Service operators can use web analytics to analyse their server and service logs regarding all sorts of statistics of online user action and activities: site visits, transactions, and use patterns. These digital trails offer advanced opportunities for dialogue between developers and users, and means to tap in and collaborate with user owned interconnected resources and services.

CONCLUSION

We have reported on user involvement methods deployment from a longitudinal case. The aim was to give a rich picture of development practices, including the emergent method repertoire of the developers. This position paper focuses on a neglected slowly changing contextual aspect: developer–user social distance. While the context of use has been in focus since the beginnings of user-centred design, the design context has been found between the lines and in the margins until recent debates [5, 13, 14, 16]. The concept of developer–user social distance brings these two, design context and use context, together. It has the potential to overcome what has been described as a 'heroic view' of design, where developers are understood in too simplistic notions of either omnipotent heroes or malevolent devils. Furthermore, the unclear role of informal engagement and personal experience in changing design and use contexts can be resolved by considering shifts in developer–user social distance.

REFERENCES

1. Bruns, A. *Blogs, Wikipedia, Second Life, and Beyond: From Production to Produsage*. Peter Lang, New York. 2008.
2. Elliot, M. and Kraemer K. Computerization Movements and Technology Diffusion: From Mainframes to Ubiquitous Computing. Information Today. 2008.
3. Estes, J., Schade, A., and Nielsen J. *Social Media User Experience*. Technical report. Nielsen Norman Group 2009.
4. Holzapfel, N. User-centred design and the user-driven web. M.A. Interactive Media. University of the Arts London. 2008.
5. Hyysalo, S. *Health Technology Development and Use*. Routledge, New York. 2010.
6. International Organization for Standardization. *Ergonomics of human-system interaction – Usability methods supporting human-centred design.* Tech. rep. ISO/TR 16982:2002.
7. Jenkins, H. *Convergence Culture: Where Old and New Media Collide*. New York University Press, New York. 2006.
8. Johnson, M. User Involvement, Social Media, and Service Evolution: The Case of Habbo. *Proc. HICSS 2010*, 1–10.
9. Johnson, M., Hyysalo, S., and Tamminen, S. The Virtuality of Virtual Worlds, or What We Can Learn from Playacting Horse Girls and Marginalized Developers. *Symbolic Interaction*, 2010, 33(4): 603–633.
10. Johnson, M. and Hyysalo, S. Lessons for Participatory Designers of Social Media: Long-Term User Involvement Strategies in Industry. *Proc. PDC 2012*, 71–80.
11. Johnson, M. How Social Media Changes User-Centred Design, D.Sc. Thesis, Aalto University, 2013.
12. Shirky, C. *Here Comes Everybody: The Power of Organizing Without Organizations.* Penguin Press, New York. 2008.
13. Stewart, J. and Williams R. "The wrong trousers? Beyond the design fallacy: Social learning and the user". *Handbook of critical information systems research: Theory and application*. Ed. by D. Howcroft and E. Trauth. Cheltenham, UK: Edward Elgar, 2005, pp. 195–221.
14. Svanæs, D. and Gulliksen J. Understanding the context of design: towards tactical user centered design. *Proc. Nordichi 2008*, pp. 353–362.
15. Tapscott, D. and Williams, A.D. *Macrowikinomics: Rebooting Business and the World*. Portfolio/Penguin, New York. 2010.
16. Woolrych, A., Hornbæk, K., Frøkjær, E., & Cockton, G. (2011). Ingredients and meals rather than recipes: A proposal for research that does not treat usability evaluation methods as indivisible wholes. *Int. J. HCI* 27(10), 940-70.

Without a Clue What Design It Leads to: Exploring and Understanding a City and Life within It by Walking

Heli Väätäjä
Tampere University of Technology
Korkeakoulunkatu 1, 33720 Tampere, Finland
heli.vaataja@tut.fi

ABSTRACT
Starting design by exploring, experiencing and understanding the city and life within it by walking amazed an HCI researcher in several ways. First, it revealed how dependent one is on structured traditional methods of user-centered design. Second, it showed how fixed one is to quickly start aiming for providing solutions to problems or trying to find needs to fulfill. Third, one easily drifted to thinking about the current available technologies and technological infrastructure in the city and how it could be used to engage people or study the city life. Fourth, it demonstrated how the characteristics of each team member affected their approach and orientation towards approaching the challenge of having no clear design goal in the first place, accepting the frustration of having to search for it and having to use considerable amount of time to explore the city and experience it with all senses and with an open mind. Accepting the frustration and giving in to the true exploring of the city, provided insights of the city life that could not have been found with traditional approaches that aim to identify needs, or by using technology-driven design goals.

BACKGROUD OF THE CASE STUDY
In this case study I describe what happened when a group of four HCI researchers was given a challenge to use new methods in exploring, experiencing and understanding the city life. Our group consisted of two researchers with the educational background in engineering, one researcher with a background in design and one with a background in architecture. We had five days to complete our design challenge ("Urban sensoria", by A. James), which was defined widely as using ubiquitous computing for personal urban sensing and to explore the use of ubiquitous systems and solutions for the creation of content, preservation, and sharing of experiences.

This paper describes what happened during this week from the point of view of one of the group members (author). Paper briefly describes the methods applied in the project, specifically highlighting the process, its phases and challenges that the group encountered during the project week. In addition, the findings of the project are briefly illustrated to give a concrete example how design ideas may emerge from study findings that seemingly do not provide explicit needs or requirements for design or technology development.

DESCRIPTION OF THE USED APPROACH
As a start for the creative process, we were given a brief on the methods to be applied in the creative process. The premise for the project was to explore the socio-cultural context with different senses by using ethnographic approach (Atkinson et al. 2007, Hammersley & Atkinson, 2007) and using these findings in the group's creative process. The goal was to aim for creating a theoretical framework based on the findings, increase cultural awareness through exploring, sensing and understanding the city, and to document, create and share the findings and solutions.

We were briefed to explore the city as a group but on the other hand independently, as the experience and sensing of the city is shaped by not only the local culture and its memory, but by each person's own prior experiences. Technological solutions, such as audio, photos, video clips, sensor data could be used to capture the city life. The idea was to bring the subjective experiences and observations of the city to group meetings and start the creative process from identifying from the individual findings the interesting themes to explore as a group. The expected outcome of the project could be new methods for exploring, new content created, creation and evaluation of new technological solutions, or new tools.

Walking as a way of exploring the city

Prior to setting off to the city, we were introduced to different types of approaches to use walking and mapping as a way to explore the city. The approaches were related psychogeography (e.g. Bassett, 2004, Pinder, 1996), and specifically dérive and algorithmic psychogeography were discussed. Dérive is "drifting", "an unplanned journey through a landscape", described as follows (Debord, 1955, translated by Knabb): "*In a dérive one or more persons during a certain period drop their usual motives for movement and action, their relations, their work and leisure activities, and let themselves be drawn by the attractions of the terrain and the encounters they find there…"*. In algorithmic psychogeography, walking is guided by an algorithm or fixed pattern (Bassett, 2004). An example is

turning right after two blocks, walking one block, turning left, walking two blocks, turning right etc.

IMPLEMENTATION OF THE APPROACH

We started off to the city and focused first on observing and the city life, urban rhythm, physical context, people's movement in the city, groups of people and their actions. We took notes, shot photos and videos and tried to identify patterns related to people's movement and locations within the city center. We also interviewed from where and why a person had come to a specific location. Figure 1 illustrates, how we observed in certain areas the flow of people and formation of pairs and groups.

Figure 1. Observing city life and patterns of people moving in selected areas.

We made several attempts to come together, discuss and identify something to design for and failing miserably. And ended up going out again time after time, to observe the city life. When coming together we drifted between technology and design trying to think about existing technological infrastructure within the city, like public screens how to engage people into using them, thinking about how to implement temperature sensing into our observations of how people use space on the streets, how to use QR codes for accessing old photos of buildings at a certain location and think about how to design something engaging people into action. We were half way through the five days, and we had nothing concrete, that we could agree on as a goal, nothing that seemed worth exploring and meaningful enough to be implemented or studied in practice. Example of one of our first attempts that we dropped is shown in Figure 2.

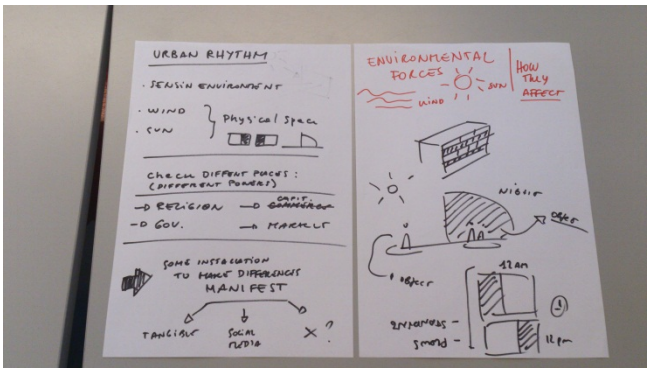


Figure 2. Exploring the rhythm of the city.

We set off once again to explore the city, frustrated, and separately. This time one of the team members got something interesting in his camera view: graffiti, stickers and posters. When coming together we again first started to think about how could we apply technology to replace these practices as they are illegal, but then we realized that by looking at these practices we might actually learn something about the local life as these practices are manifestations within their context. At this point we finally were ready to break free from the technology driven design and took the map of the city center, divided it into four equally sized blocks and map these areas through these three types of media with GPS enabled mobile phones capturing photos of the selected media types – each poster, sticker and graffiti that we came across within the area that each of us was assigned.

Mapping the city through urban traces

As we set out to explore and captured altogether 512 geotagged photos from the city center of the three types of media (stickers, posters and graffiti) we were looking into the question of why in the era of digital technology, people still use the non-digital and illegal forms of communication, what can we learn about the city life and the practice itself?

To map the captured photos, we decided to upload them to Flickr (see Figure 3). We used both qualitative and quantitative analysis of the photos. We used in the qualitative analysis three categorizations based on 1) the type of media (graffiti, sticker or poster), 2) the surface on which the media was found (poles, trash bins etc.), and 3) the content or theme of the media (music, politics, sports etc.). We used these categories and codes within the categories to tag each photo (see Figure 4). Quantitative analysis was used to reveal relationships between media, theme and surface.

Figure 3. Uploaded geotagged photos to Flickr.

Figure 4. Examples of categories and codes in the analysis.

THE WHOLE PROCESS AND THE OUTCOME

As a whole, the process that we went through during the week is illustrated in Figure 5. The figure illustrates, how we went through multiple phases of exploring, coming finally up with the idea, and implementing it, and as both the exploration by photographing as well as actual analysis of the collected data emerged, we gained understanding of the city life, practices related to these media, and found a new map of the city, that was created by the captured and mapped photos and dependent on the media (see Figure 6).

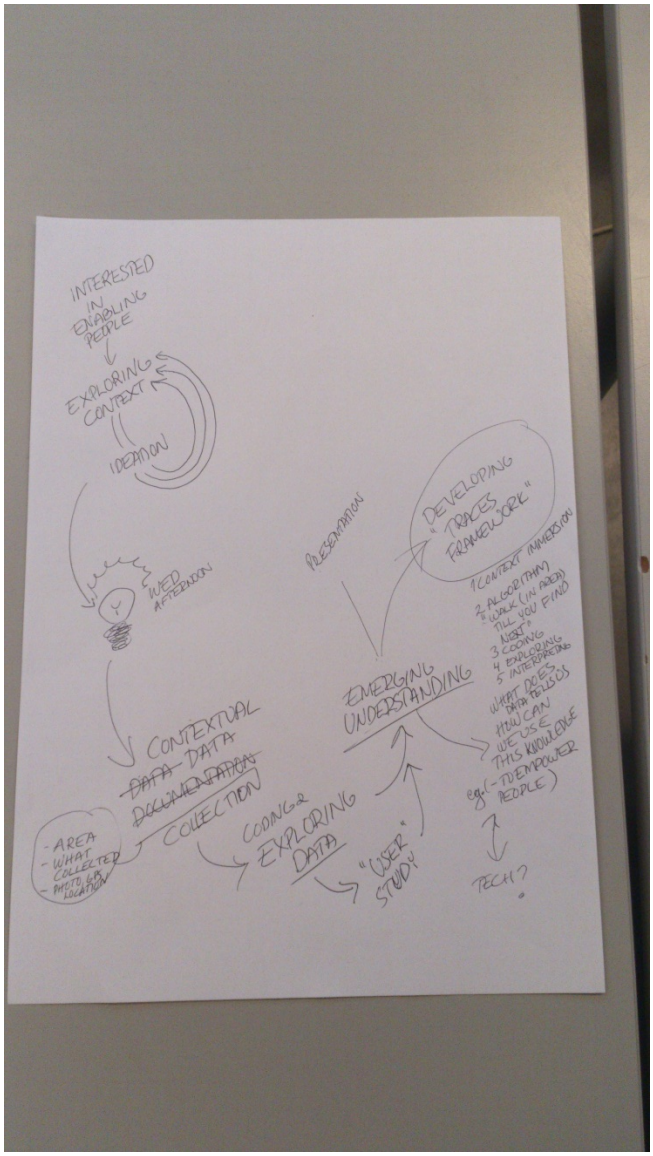


Figure 5. The method that lead to successful outcome.

WHY WE FINALLY SUCCEEDED?

We used considerable amount of time to struggle with our sticky habits of HCI related practices, methods that we had used in our prior studies for eliciting user needs and requirements or for innovation, own personal preferences, and trying too hard and too fast to get into the design and implementation phase. Specifically, each of us also was stuck with the basic university education background in one way or the other – either engineering, architecture or design, which was manifested in how we approached trying to identify either implementation possibilities or what to design for. It therefore took quite some time break free from our background and habits, and prior approaches that we had each found successful in our own prior research when designing and developing solutions for relatively clear goals and needs.

The succeeding in the challenge and starting truly to explore the city can be attributed to the multidisciplinary background of the team as it enabled to

- identify the topic for the focus of the study, as one of the team members had photography as a hobby and he therefore had an eye to look for topics to photograph,
- the innovative way of using technology and its affordances in capturing and mapping the city life, as the team members were used to think about utilizing affordances of available services and technological enablers in their research in data collection and visualization,
- the methodological approaches in exploring the city through walking, that the team applied in the task and
- the qualitative and quantitative analysis approaches used in the analysis of the collected photos that were based on team members' prior approaches used in analysis of data and applied for the analysis and coding of the geotagged photos.

Although we did not implement in our project a new technological solution to replace an existing practice or augmenting the reality with digital information, our solution to capture photos and share them in Flickr documents and captures the reality of that moment and part of history that may be cleaned up tomorrow. In addition, by exploring this type of existing practices in its specific context, we may find new requirements for supporting the practices with technological solutions as well.

ACKNOWLEDGEMENTS

I thank our design team members Anna Luusua, Lorenzo Davoli and Shahram Jalaliniya for sharing the unique and memorable experience that each of us experienced in our own way. The research of the author was funded by the Finnish Doctoral Program on User-Centered Information Technology.

REFERENCES

[1] Atkinson, P. Coffey, A., Delamont, S. Lofland, J. Lofland, L. Eds. (2007). Handbook of Ethnography. SAGE.

[2] Bassett, K. (2004). Walking as an aesthetic practice and a critical tool: some psychogeographic experiments, Journal of Geography in Higher Education, 28,3, 397-410.

[3] Brewer, J., & Dourish, P. (2008). Storied spaces: Cultural accounts of mobility, technology, and environmental knowing. *International Journal of Human-Computer Studies*, *66*(12), 963-976.

[4] Dourish, P., Anderson, K., & Nafus, D. (2007). Cultural mobilities: Diversity and agency in urban computing. *Human-Computer Interaction–INTERACT 2007*, 100-113.

[5] Hammersley, M., Atkinson, P. (2007). Ethnography: Principle in practice. 3rd Ed. Routledge.

[6] Pinder, D. (1996). Subverting cartography: the situationists and maps of the city. Environment and planning A, 28, 405-427.

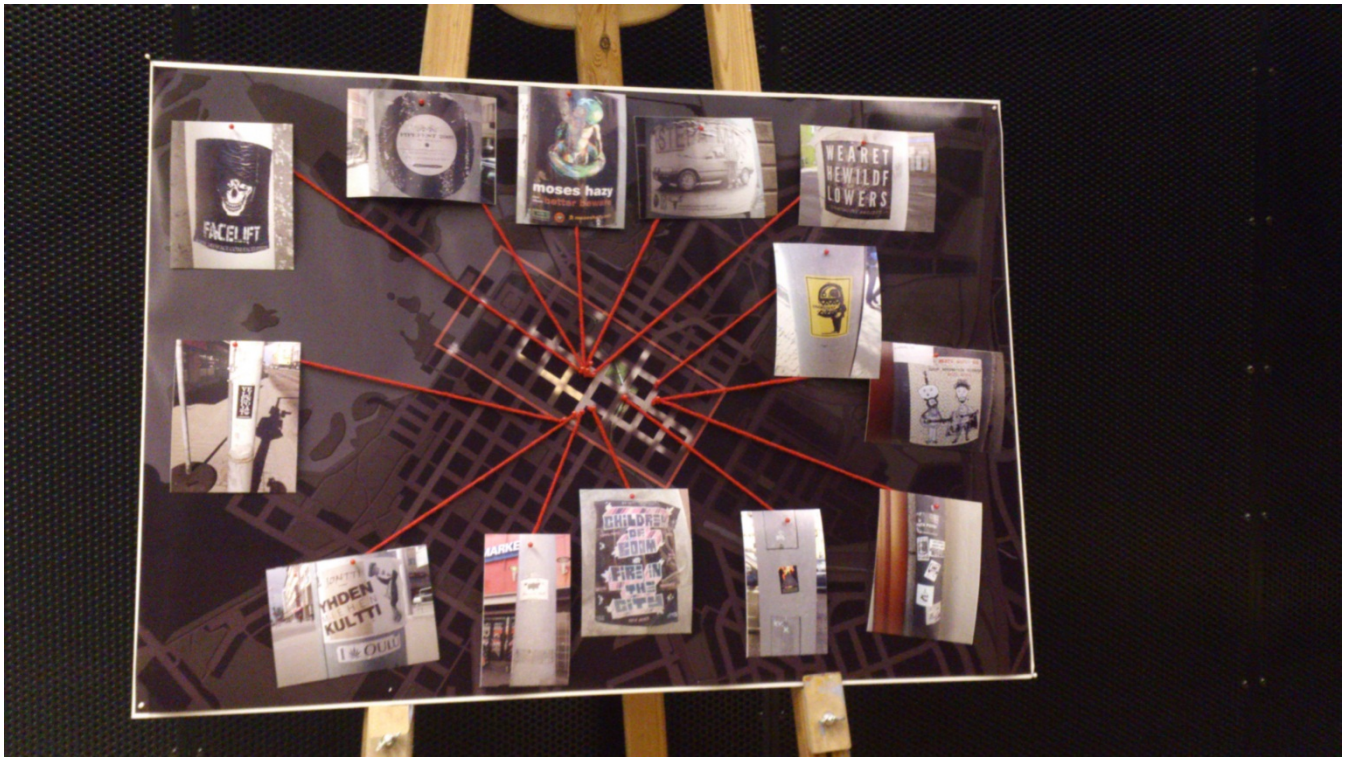


Figure 6. The invisible edges of the city center lightened up by stickers on poles.

European Science Foundation provides and manages the scientific and technical secretariat for COST

COST is supported by the EU RTD Framework Programme