CASSM (*formerly OSM*) for Usability Evaluation and Design

EPSRC GR/R39108/02

Introduction

This project aimed to develop and test a novel approach to usability evaluation focusing on conceptual structures rather than, as is usual, users' tasks. We have achieved all the stated objectives and more, ending with a tested methodology and a complete, coherent, downloadable package for dissemination purposes. As the work has developed we have changed our perspective on it, resulting in a new name: Concept-based Analysis of Surface and Structural Misfits (CASSM); throughout this final report we refer to it by this new name.

Background/Context

Two kinds of approach have dominated analytical work on the usability of interactive systems: *heuristic* (or *checklist-based*) approaches giving a swift assessment of look-and-feel (usually without consideration of the tasks the system is designed to support), such as Heuristic Evaluation (Nielsen, 1994); and *procedure-based* approaches for assessing the difficulty of each step of typical user tasks, such as Cognitive Walkthrough (Wharton *et al*, 1994).

CASSM presents a third approach, namely the analysis of *conceptual misfits* between the way the user thinks and the representation implemented within the system. This approach has been shown, through this project, to fill a useful niche in the space of analytical evaluation techniques.

Within HCI, we are not aware of any other approach that supports the systematic analysis of the conceptual fit between user and system. Moran's (1983) ETIT and Payne's Yoked State Spaces (YSS: Payne *et al,* 1990) strongly informed our thinking about CASSM, as did our own earlier work on PUM (Blandford & Young, 1996) and ERMIA (Green & Benyon, 1996); neither ETIT nor YSS was codified in a form that would make it possible for non-specialists to apply those approaches. While PUM and ERMIA were both codified (in tutorial materials), we perceive both as demanding too much precision and detail to apply in routine practice. The final approach that has heavily influenced our thinking is Green's Cognitive Dimensions of Notations (CDs: Green, 1989; Blackwell & Green, 2003). CDs provide a very informal vocabulary for various misfits between user and system. We have developed precise definitions of many CDs, which capture their meanings reasonably faithfully and are now encapsulated within the Cassata analysis tool described below.

CASSM focuses on the quality of fit between the user's conceptual model and that implemented within a system, encompassing a method for both data gathering and data analysis. Within an evolutionary development lifecycle, a recognition of poor fit should highlight opportunities for re-design (Hsi, 2004).

CASSM and Cassata: a brief introduction to the final versions

CASSM is a method and modelling representation; CASSM analysis is supported by a prototype tool: Cassata. CASSM compares the users' concepts (identified by an appropriate data gathering technique such as interviews, think-aloud protocols or Contextual Inquiry) with the concepts implemented within the system and interface (identified by reference to sources such as system documentation or an existing implementation). Conceptual analysis draws out commonalities across similar users to create the profile of a typical user of a particular type; the analyst can then assess the quality of fit between user and system, notably *surface misfits*: concepts salient to the user but partially or wholly missing from the system, and concepts imposed on the user by the system. As analysis proceeds, the analyst will start to distinguish between entities and attributes (as defined below), and to consider what actions the user can take to change the state of the system. Finally, for a thorough analysis, various relationships between concepts are enumerated to identify *structural misfits*, cases where the relationships between concepts in the user's model do not coincide with those in the system, as described more fully below.

Surface misfits go back, although not by that name, to Norman's (1986) analysis of whether the user of the system could understand the designer's conceptual model and whether the system appropriately implemented pre-existing user concepts. The advance made by CASSM is to provide a usable methodology with which to identify them. Structural misfits, in contrast, are hardly present in previous usability evaluation methods, so we shall describe them at slightly more length.

Figure 1 shows a screen shot from Cassata, the tool we developed to facilitate the description and analysis of CASSM models (paper-based representations can also be used). This particular example is a partial description of a word processor document. Such documents contain numbered sequences—of figures, tables, diagrams, etc—and the hypothetical user is taken to have a clear concept of *number-sequence* as an attribute of the set of figures, denoted *set-of-figs*. Each *figure* has an attribute which is its particular *number*, and changing a figure number changes the overall sequence of figure numbers.

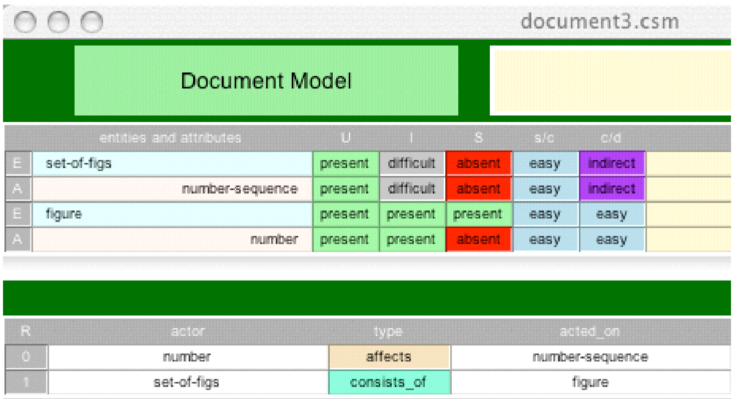

Fig. 1. Cassata data table for a partial description of a document. The upper table describes concepts (i.e. entities and their attributes); the lower describes relationships between those concepts.

The top half of the window shows information about concepts (entities such as *figure* and attributes such as *number*): for each concept, three columns show whether it is present, difficult or absent for the user, interface and system respectively; the next two columns show how easy it is to set or change the value of an attribute, or to create or delete an entity; the final column is a notes area in which the analyst can add comments.

To take the first row as an example: the `set-of-figs` is a conceptual entity that is meaningful to the user, is not clearly represented at the interface ('difficult') and is absent from the underlying system model. It is easy to create a set of figures (because this happens automatically as the user adds figures) but it is deleted indirectly (by deleting *all* the individual figures). The bottom half of the window shows information about relationships (such as *affects* and *consists_of*) between concepts. In the example, the two lines of input state that changing any *number* (of a *figure*) affects the *number-sequence* (of the *set-of-figs*) and that a *set-of-figs* consists of *figures*. Automated analysis (within Cassata) of this particular representation would show that the system is prone to *repetition viscosity* – that deleting or inserting a figure results in the user having to perform a repetitious sequence of actions to restore the figure numbers to incrementing as required.

Key Advances and Supporting Methodology

The following key conceptual advances were made by this project.

- **Surface and structural misfits**: This distinction has not previously been made. Previous evaluation methods have rarely delved below the surface of the system. Our analyses of several types of system have shown that some systems, such as intranets, are relatively unchallenging to analyse, because any misfits they contain are on the surface. These systems typically do not exhibit the difficulties encapsulated within Green's CDs. In contrast, systems that enable the user to manipulate structures, such as word processors and drawing tools, can fruitfully be analysed in terms of CDs, and may suffer from structural misfits as well as surface ones. Relationships between concepts only become interesting in the case of structural misfits. This insight prompted the name change from OSM (which emphasised the ontology without mentioning the structure) to CASSM.
- **Avoidance of task analysis**: Initially the CASSM representation included explicit description of user actions and a form of task analysis, breaking a conceptual action (an action that has domain significance to the user) into a sequence of device actions. This approach precluded the automatic analysis of CDs and also made analysis more time-consuming and pettifogging than we would have liked. Eventually, iterative refinement and testing led to a representation where the important features of user actions can be described without explicitly naming those actions.
- **Creator–Composer–Consumer chains**: We have identified an important case of indirect concept representation, where a software tool is created by one person, the creator; software is composed by a second person, the composer; and that software is used by a third person, the consumer. For successful operation with few misfits, the creator of the original tool needs to understand the concepts being brought to an interaction by both the composer (the immediate user of their system)

and the subsequent consumer of the resulting system. This emerged while working with Cancer Research UK on evaluating Tallis, a knowledge representation tool that allows clinicians to compose clinical guidelines for patient treatment. In this case, it became apparent that the design of the tool itself heavily influences the design of the interaction for the end user of those guidelines (Connell *et al*, submitted). CASSM has helped with comparing and reasoning about the models of the different user groups in such chains.

Having highlighted the most important conceptual advances, we now present an account of the project's achievements against the initial objectives, and additional achievements that have extended the project beyond the work anticipated.

Objective 1: *Testing CASSM thoroughly, and adapting it if any areas for improvement are identified:*

The approach has been tested on a variety of examples, intentionally covering a broad range of system types. Five examples are included on the project web site, ranging from a hospital intranet and two ticket machines (both relatively simple 'walk up and use' systems) to the complex Tallis knowledge representation tool. Further analyses (of a robotic arm, a heating controller and the London Ambulance (LAS) control system) are included in the CASSM tutorial. The LAS analysis is particularly important because it illustrates the application of CASSM to a complex socio-technical system (Blandford, Wong, Connell & Green, 2002).

Many of the examples listed above were initially developed using an earlier version of CASSM, but all have been re-worked in the current version, increasing our confidence in the appropriateness of the current representation and the usability of the Cassata tool.

Objective 2: *Testing our hypotheses that CASSM is better suited than existing HCI techniques for reasoning about devices that support ill-structured tasks and usability in context, for detecting 'misfits' and for supporting reuse.*

We have conducted three systematic studies that compare CASSM to other established HCI techniques. Connell *et al* (2004) compare a CASSM analysis and a Cognitive Walkthrough (Wharton *et al*, 1994) of two ticket vending machines against observational and interview data of users of those machines in context (i.e. at London Underground stations). There, it was found that the overlaps between the three approaches were surprisingly small (i.e. many observed user difficulties were not picked up by either analytical technique, and the two analytical techniques focused on different aspects of interaction). Further analysis showed that many of the issues highlighted, particularly by CASSM, were ones that would not be directly observable, in that they concerned the user's conceptual understanding which the machines prevent the user from expressing directly. Blandford *et al* (2004b) compare the experiences of conducting a Heuristic Evaluation, a Cognitive Walkthrough, a Claims Analysis and a CASSM analysis of digital libraries; we found that Heuristic Evaluation and Cognitive Walkthrough only address superficial aspects of interface design (but are good for that), whereas Claims Analysis and CASSM can help identify deeper conceptual difficulties (but demand greater skill of the analyst); however, none fit seamlessly with existing digital library development practices.

The final, most thorough, comparative analysis is a major review of evaluation methods. Blandford *et al* (submitted) compare seven different analytical UEMs against each other and against video data for a robotic arm. This employed a very rigorous approach involving rational reanalysis, identifying commonalities and contrasts between techniques and providing accounts of *why* a particular technique yielded the insights it did. Each analytical technique was found to focus attention on just one or two out of five categories of issues. Accounts of each technique's scope included whether each issue was found, should have been found but was not, should not have been found but was (by presumed craft skill), or was not found by the technique. CASSM, appropriately, was found to identify important issues regarding conceptual fit that were missed by other techniques.

Turning to reuse, we have conducted two studies that explicitly address this question. The first was an analysis of the LAS control system. In this context, the same computer system is used by four different user groups (call takers, telephone dispatchers, radio operators and allocators); while the call takers have a substantially different interface to the system, the remaining three user groups work with almost identical interfaces. Thus, it was possible to compare the (single) system and interface design against the conceptual models of the different user groups. The second study focused on a single user description and multiple systems – in this case, scheduling systems. In CASSM, reuse of the system description is simple and straightforward. That is certainly not true of process or heuristic based analytical techniques.

As noted above, the Creator–Composer–Consumer framework for thinking about system design also makes use of different system and user oriented descriptions. While this is not a case of simple re-use, the ways in which the creator/composer perspective can 'leak through' into the consumers' system can be studied by considering how much of the composer model is reused for the consumer model.

Objective 3*: Creating a library of worked examples.*

We have created a library of documented worked examples, available from the CASSM website (http://www.uclic.ucl.ac.uk/annb/CASSM/). In addition, three further examples are included in the tutorial materials for illustrating the CASSM approach and the use of the Cassata data analysis tool.

Objective 4*: Producing formal definitions of a range of misfits (including many of Green's Cognitive Dimensions).*

We have produced formal definitions of a subset of Green's CDs, as reported by Blandford *et al* (in press). We have systematically worked through the set of CDs, considering which can and cannot be formalised using the CASSM representation. Some, such as *necessary device abstractions* and *visibility* can often be detected as surface misfits. Others, notably *viscosity*, *premature commitment* and *hidden dependencies*, demand structural analysis, now implemented within Cassata.

Objective 5*: Developing and testing a prototype demonstrator tool to support CASSM analysis.*

We have developed, tested and documented the Cassata tool mentioned above. Cassata is a second-generation tool. The first design, called OSMosis, turned out to be unexpectedly hard to use, and did not support the automatic identification of CDs. Rethinking the tool was a salutary experience that forced us to reflect on the fundamentals of the CASSM ontology and analysis process.

Cassata is implemented in the cross-platform language Tcl/Tk, and runs on both PCs and Macs (probably also on Unix boxes but not yet tested). It should be regarded as a proof of concept demonstrator; it is certainly not an industry standard robust tool. Nevertheless, not only did we find Cassata invaluable for the many analyses performed during this project, but we also regard it as a key component of the dissemination package.

Objective 6*: Presenting accounts of the usability of medical information systems, expressed in CASSM terms.*

We have analysed a hospital intranet, a medical decision support system and an ambulance dispatch system in CASSM terms. Although the hospital intranet system analysis was difficult for organisational reasons, because we were required to obtain ethical clearance and then has some difficulty in recruiting users with much experience of the intranet, it proved invaluable in highlighting the centrality of user data in constructing models and was a core exemplar in devising the distinction between surface and structural misfits. The work with Cancer Research UK was also slightly hampered by there being few established users of Tallis, but this study helped focus our ideas around creators, composers and consumers.

As well as all the stated objectives, we have achieved the following:

New objective 7*: Developing and delivering a tutorial on CASSM.*

We have developed and delivered a tutorial on CASSM. This tutorial has been tested in-house (with students on the MSc in HCI with Ergonomics at UCL) in each year of the project, and was also presented at Interact 2003. The final version of the tutorial is available from the project web site. The tutorial and the Cassata tool have been distributed to HCI specialists in other universities, including Sheffield Hallam (from whom we have already received useful feedback), the Open University and Carnegie Mellon. Roast *et al* (2004) and Mann (2004) include brief descriptions of their uses of CASSM within their own research work. Since both made use of interim versions of the approach, and neither was conducting a formal evaluation of CASSM, their reports do not explicitly focus on the usability or utility of the approach. Nevertheless, the fact that there are already reports of the use of CASSM by researchers who have no direct connection with the project team is highly encouraging.

New objective 8: *Initiating knowledge transfer to industry.*

UCL has been awarded an Innovation Grant of £10K from the London Development Agency, in collaboration with Radiant (www.radiant-digital.com), to investigate the feasibility of adapting and packaging CASSM specifically for the evaluation of e-commerce sites. The details of this work are currently commercially sensitive. In addition, we are currently in negotiation with a large information supplier on applying CASSM to new systems that are under development and have also received an outline expression of interest from a company in the telecommunications sector.

Project Plan Review

The work was slightly disrupted early on by Blandford's move from Middlesex University to UCL, but the effects of the move were minimised by the support and cooperation of both institutions, and of the project RA (Iain Connell). There were no actual deviations from the project plan.

Research Impact and Benefits to Society

All beneficiaries listed in the original proposal remain valid. As shown by our various comparative and scoping analyses of CASSM, it is an approach that fills an important niche in the space of usability evaluation methods. It is the only method that we know of that explicitly integrates the collection of user data with an analytical approach. We have aimed throughout the project to create a method that is usable, useful and used. Within the project, we have focused our attention on the conceptual development of the approach and on encapsulating it in a way that facilitates knowledge transfer. We have tested its usefulness, particularly through the comparative studies outlined above and through our interactions with developers at CRUK. It would be unrealistic to expect to see substantial take-up of a new method within the lifetime of a project such as this (the approach has only stabilised within the past year or so), but the indicators are highly positive, as discussed above.

The benefits to society of this work will come about in two ways. One is through direct application of the approach to help in the evaluation of existing systems and the subsequent re-design of systems to reduce the misfits between user and system. The second is indirect. HCI is feeling its way towards evaluation techniques that are effective but not too costly, too hard to learn, nor too ill-defined. CASSM, we believe, steers an excellent compromise between these rocks and whirlpools: reasonably cheap in both time and expense, easy to learn, adequately precise, and able to suggest positive design improvements. We hope that the CASSM approach will itself be taken up and adopted within other research and development work. Even if other approaches are found to be more successful, we would like to think that CASSM has set a new standard for combining usefulness and usability in an evaluation method, leading to the development of systems that better fit the user's work.

Explanation of Expenditure

There were no deviations from the expected expenditure. With the agreement of EPSRC, we employed a programmer for a very short period to implement OSMosis (the first CASSM analysis tool); this was paid for under the appropriate budget heading. Connell was supported in attending conferences and tutorials as part of his research training, and regularly gave presentations, internal and external, on the project. He is currently seeking a suitable lectureship.

The work with CRUK proceeded as planned. It is impossible to put a precise cash value on their contribution to the project, but the collaboration was of mutual benefit (see attached letter from Dr Rose).

Further Research or Dissemination Activities

Immediate industrial dissemination activities have been described above. We are also exploring the possibility of presenting the tutorial as an industrial short course. We are preparing two further academic papers on the work, one on reuse of representation as described above; the other developing further the socio-technical analysis based on work at LAS.

We have intentionally closed the project off at an effective breakpoint with few loose ends, presenting clearly defined deliverables that package up the approach. However, some important questions have emerged through this work that will be pursued when we have had time to fully reflect on the findings of this project. These include:

- Further investigation of the usability and utility of the approach, through further knowledge transfer activities. The planned work with Radiant will initiate this.
- Further investigations in the medical domain, studying how clinicians conceptualise key parts of their day-to-day work (such as applying clinical guidelines) and how that understanding can inform design.

- Investigation of the misfits in user interactions with novel technologies such as haptic, gestural VR or augmented reality systems and whether CASSM can provide useful leverage in reasoning about such interactions (going beyond the concepts that are easily represented through classic GUI systems).

Project publications

Blandford, A. & Connell, I. (2003) Ontological Sketch Modelling (OSM): Concept-based Usability Analysis *Proc. Interact 2003*. 1021-1022. IOS Press.

Blandford, A., Connell, I. & Green, T.R.G. (2004a) *Concept-based Analysis of Surface and Structural Misfits (CASSM) Tutorial notes*. CASSM working paper available from http://www.uclic.ucl.ac.uk/annb/CASSM/

Blandford, A. & Green, T.R.G. (2001) From tasks to conceptual structures: misfit analysis. In J. Vanderdonckt, A. Blandford & A. Derycke (Eds.) *IHM-HCI2001 Vol. II* (short paper). Cépaduès-Éditions. 113-116.

Blandford, A., Green, T.R.G. & Connell, I. (in press) Formalising an understanding of user–system misfits. To appear in *Proc. EHCI-DSVIS 2004*. Springer: LNCS.

Blandford, A., Hyde, J., Connell, I. & Green, T.R.G. (submitted) Scoping analytical usability evaluation methods: a case study. Submitted to *HCI Journal*.

Blandford, A., Keith, S., Connell, I. & Edwards, H. (2004b) Analytical usability evaluation for Digital Libraries: a case study. In *Proc. ACM/IEEE Joint Conference on Digital Libraries*. 27-36. (30% CASSM, 70% GR/N37858)

Blandford, A. E., Wong, B. L. W., Connell, I. & Green, T. R. G. (2002) Multiple viewpoints on computer supported team work: a case study on ambulance dispatch. In X. Faulkner, J. Finlay & F. Détienne (Eds.) *People and Computers XVI: Proc. HCI'02*. 139-156. Springer.

Connell, I., Blandford, A. & Green, T.R.G. (2004) CASSM and cognitive walkthrough: usability issues with ticket vending machines. *Behaviour and Information Technology*, 23(5). 307-320.

Connell, I., Blandford, A., Rose, A. & Green, T. R. G. (submitted) Knowledge representation environments: an investigation of the CASSMs between creators, composers and consumers. Submitted to *International Journal of Human–Computer Studies*.

Connell, I., Green, T. & Blandford, A. (2003) Ontological Sketch Models: highlighting user-system misfits. In E. O'Neill, P. Palanque & P. Johnson (Eds.) *People and Computers XVII, Proc. HCI'03*. 163-178. Springer.

Green, T. R. G. & Blandford, A. E. (2004) Downloadable package: CASSM tutorial, worked examples, Cassata tool and supporting documentation. From http://www.uclic.ucl.ac.uk/annb/CASSM/CASSMdownloads.html

Other references

Blackwell, A. & Green, T. R. G. (2003) Notational systems – the Cognitive Dimensions of Notations framework. In J. Carroll (ed.), *HCI Models, Theories and Frameworks*, Morgan Kaufmann. 103-134.

Blandford, A. E. & Young, R. M. (1996) Specifying user knowledge for the design of interactive systems. *Software Engineering Journal*. 11.6, 323-333.

Roast, C., Dearden, A. & Khazaei, B. (2004) Enhancing Contextual Analysis to Support the Design of Development Tools. In S. Fincher, P. Markopoulos, D. Moore & R. Ruddle (Eds.) *People and Computers XVIII.* Springer. 297-313.

Green, T. R. G. & Benyon, D. (1996) The skull beneath the skin: entity-relationship models of information artifacts. *International Journal of Human-Computer Studies*, 44, 801-828

Green, T. R. G. (1989) Cognitive dimensions of notations. In A. Sutcliffe and L. Macaulay (Eds.) *People and Computers V*. Cambridge University Press. 443-460

Hsi, I. (2004) Measuring the conceptual fitness of a computing application in a computing ecosystem. *Proc. WISER 2004 (ACM Workshop on Interdisciplinary Software Engineering Research)*, November 5, 27-36.

Mann, P. (2004) Design for design: Requirements for graphical communication in computer-supported collaborative work (CSCW) in design. In A. Dearden & L. Watts (Eds) *Proceedings of HCI 2004 (Vol. 2).* Research Press. 229-230.

Moran, T. P. (1983) Getting into a system: external-internal task mapping analysis. In A. Janda (ed.), *Human Factors in Computing Systems*, 45-49.

Nielsen, J. (1994) Heuristic evaluation. In J. Nielsen & R. Mack (Eds.), *Usability Inspection Methods*. New York: John Wiley. 25-62.

Norman, D. A. (1986) Cognitive engineering. In D. A. Norman and S. W. Draper, Eds. *User Centered System Design,* Hillsdale NJ: Lawrence Erlbaum, 31-62

Payne, S. J., Squibb, H. R. & Howes, A. (1990) The nature of device models: the yoked state space hypothesis, and some experiments with text editors. *Human-Computer Interaction*, 5. 415-444.

Wharton, C., Rieman, J., Lewis, C. & Polson, P. (1994) The cognitive walkthrough method: A practitioner's guide. In J. Nielsen & R. Mack (Eds.), *Usability inspection methods* (pp. 105-140) New York: John Wiley.