INTERNATIONAL DESIGN CONFERENCE - DESIGN 2004
Dubrovnik, May 18 - 21, 2004.

STUDIES OF INTUITIVE INTERACTION EMPLOYING OBSERVATION AND CONCURRENT PROTOCOL

A. Blackler, V. Popovic and D. Mahar

Keywords: intuitive use, human factors, interface design, observational analysis, product design

1. Introduction

Recent research has explored how intuition works during product use and how products can be designed to make them more intuitive to use. Results of the research have been reported in detail elsewhere [Blackler, Popovic and Mahar 2003a, Blackler, Popovic and Mahar 2003b]. This paper is concerned with the methods used to find answers to these complex questions. This is done by focusing on the methods and results of Blackler, Popovic and Mahar [2003a] and particularly Blackler, Popovic and Mahar [2003b].

Much research suggests that intuition relies on experiential knowledge [Bastick 1982, Fischbein 1987, King and Clark 2002, Noddings and Shore 1984, Agor 1986, Bowers, Regehr, Balthazard and Parker 1990, Dreyfus, Dreyfus and Athanasiou 1986, Klein 1998]. Therefore, products that people use intuitively should be those with features that they have encountered before [Blackler, Popovic and Mahar 2003a]. Several researchers also agree that intuition is a process during which the reasoning process is not consciously available as it takes place outside the conscious mind so that the details of processing are not known [Bastick 1982, Agor 1986, Dreyfus, Dreyfus and Athanasiou 1986, Klein 1998]. The following definitions have been developed from the available research into intuition:

- Intuition is a cognitive process that is often unconscious and draws on experiential knowledge.
- Intuitive use of products involves utilising knowledge gained through other products or experience(s).

The nature of intuition presents certain problems with its study. Few experiments have been conducted specifically targeting intuition [Bastick 1982], so there is no established method to follow. When looking at intuitive interaction with products, the study of the use of individual product features rather than the product as a whole is important because it is the features that users have relevant past experience with rather than the whole product. This is especially relevant therefore when so many new products and product types are appearing on the market.

As product features were the focus of this research, looking at each feature use individually was the only way that this issue could be studied. Therefore, users needed to be observed in great detail. The challenge was to find ways of recording and coding the observations so that this level of detail could be extracted from user tests. These issues were addressed by using the Noldus Observer VideoPro software to analyse video data of observations in conjunction with a concurrent protocol.

2. Techniques

2.1 Observation

Baber and Stanton [1996] discussed the value of observation as a technique for usability evaluation. They believe that "observation can provide illuminating insight into the difficulties people face when using products." However, observation alone often does not provide enough data to draw meaningful conclusions. Therefore, observation of real people using a real product is the only way to understand if the features can be used intuitively, but a verbal protocol is needed to get at the unobservable information.

2.2 Verbal Protocols

The protocol method is widely accepted in the research community. It yields data which are unstructured and very rich, and flexible analytical methods can be used. There are four basic types of verbal protocol.

Retrospective protocol involves the users recounting what they did immediately following the test session. However, time and intervening activities can lead to decline in the accuracy of a retrospective report. Information heeded during performance of a task is the information that is in Long Term Memory and therefore that is what is reported [Ericsson and Simon 1984]. A non-conscious process (like intuition) may not be available for recall [Ericsson and Simon 1984, Wickens, Gordon and Liu 1998, Kleinmuntz 1987]. so this method would be inappropriate in this case.

'Think aloud' procedure, or Concurrent Verbal Protocol, is concurrent with the user's actions, and so should overcome the problems of participants forgetting what they had been thinking that are inherent in retrospective protocol. Kleinmuntz [1987] recommends unprompted concurrent protocol to avoid users giving inaccurate reports in retrospect, especially when prompted and therefore subject to suggestion or under pressure. The think-aloud protocol has been found to be very useful for studying interface design [Jorgensen 1990].

Co-discovery (or iterative protocol) involves having two users conducting a test and encouraging them to talk to each other as they use the product. This method is not appropriate when exploring an individual's past experience of a feature, as interaction with another person could prevent a participant from revealing their own relevant past experience of the product features, and using their own intuition.

Active intervention involves a member of the test team sitting in the room and probing the user's understanding of the product to gain a picture of their thought processes and mental models. The researchers will get more information than through observing and listening for the users to think aloud. However, this method may influence the answers given by participants [Kleinmuntz 1987] and also mask the use of intuition.

For the research reported by Blackler, Popovic and Mahar [2003a] and Blackler, Popovic and Mahar [2003b] participants were delivering concurrent protocol. This protocol method was chosen because it eliminates the problems involved with people forgetting the details when using retrospective protocol while still allowing the user to access their own past experience in order to use the product.

Intuition utilises memories and learning without the conscious mind being aware of it [Bastick 1982, Fischbein 1987, Noddings and Shore 1984, Agor 1986]. In a concurrent protocol this process is conspicuous by absence of detail and gaps in the logical steps of the commentary, because the commentary is generated in the conscious mind, which does not have access to the intuitive process. So the less reasoning is evident in the protocol, the more likely it is that intuition was being utilised.

2.3 Other Methods Employed

Structured or standardised interviews are useful if information from a number of respondents is to be combined and compared and can generate quantitative data, as the structure provides consistency [Sommer and Sommer 1997]. Interviews, although less objective, can sometimes supply more data than other methods through the interviewer's ability to explore issues in more depth, follow up ambiguities and observe body language [Davis 1996]. Questions can be tailored to be understandable to a wider range of subjects and are particularly good for exploration of complex topics [Sommer and Sommer 1997]. Questionnaires can be used to collect qualitative and/or quantitative data by employing more or less open-ended questions. Interviews are more time consuming to perform and analyse than questionnaires, but can yield richer results. A combination of structured interviews and questionnaires was used for this research.

3. Specialist software

The Noldus Observer VideoPro is a complete manual event recorder for collecting, managing and analysing observational data. It captures a level of detail not possible in live situations and that cannot be analysed easily without an automated system. The program consists of three stages - configuration, observation and analysis.

3.1 Configuration

This is the set up stage and must be completed before any data can be entered or analysed. Each behavioural class can contain many behaviours, and can have up to two modifier classes attached to it. This allows analysis of complex situations such as intuitive interaction. For example, for the behavioural class “features”, there are 21 behaviours (which are simply uses of the features named) and two modifiers (correctness of use and amount of reasoning). A behaviour can be an event or a state. An event (such as one use of one feature) may take a second or two, while a state (such as performing operation one) continues for a longer period of time. Events can occur within states.

3.2 Observation

The observation module is where the coding of the raw audiovisual data takes place. This can be stopped, started and repeated as necessary. The program allows for one or two letter or number codes to be used, but displays a name of up to 16 characters. This makes it fast to enter the codes but saves the operator from having to remember them after the coding is done. Figure 1 shows the Observer observation module during coding.

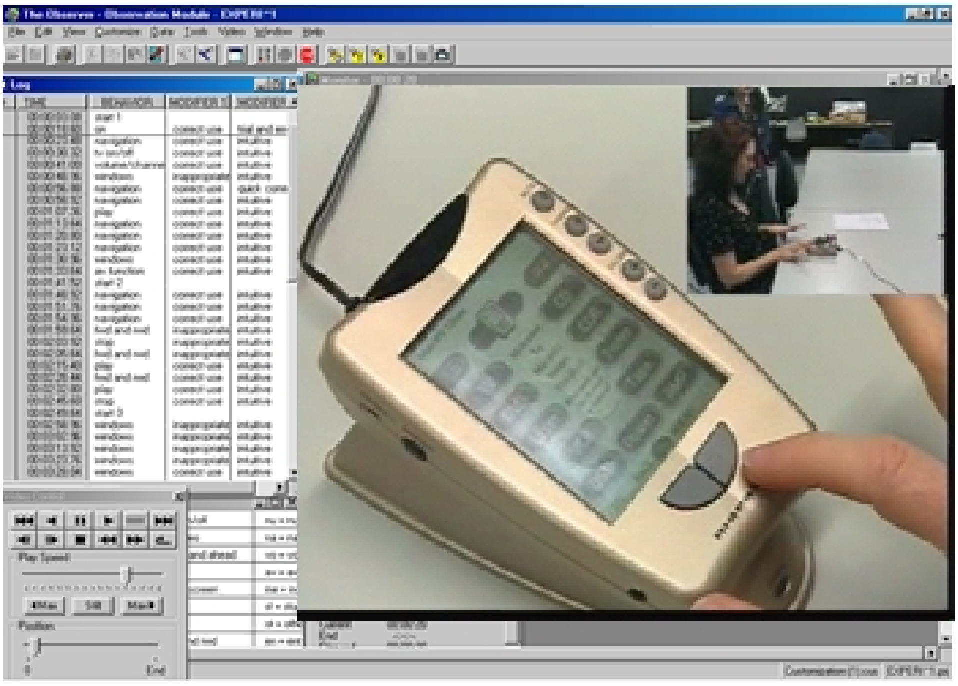

Figure 1. Observer observation module

3.3 Analysis

Observer can generate time-event logs and plots, rates, latencies and durations, for individual observations or by Independent Variable. This data can then be easily exported into packages such as Excel and SPSS for statistical analysis.

4. Experimental Method

4.1 Camera Experiments

The first set of experiments designed to investigate intuitive use of products [Blackler, Popovic and Mahar 2003a] was designed to test the thesis that intuitive use of products involves utilising knowledge gained through other products or experience(s). Based on the understanding of intuition explained above, intuition was operationalised as relevant past experience. The independent variable was level of expertise – categorised as intermediate, novice and naïve on the basis of length of experience with digital cameras. Twenty participants were asked to complete two operations with a digital camera. Immediately after finishing the operations, the participants were interviewed to establish how familiar each feature was to them. They then completed a technology familiarity questionnaire.

The technology familiarity questionnaire was designed to discover whether relevant past experience is transferable between products. The participants were asked about whether and how often they used certain consumer electronic products that had similar features to the camera, and how much of the functionality of those products they used. This questionnaire was used to calculate the Technology Familiarity (TF) score for each participant. During analysis, the time taken to complete the operations, the number of correct, inappropriate and incorrect use of camera features, and level of conscious reasoning required for each use were coded, which revealed the number of uses of each feature that were intuitive.

The results suggested that prior experience with products employing similar features helped participants to complete the operations more quickly and intuitively. The camera transferred features from other digital products, so novices with digital cameras who had experience with the features employed in the camera from using other products completed the tasks more quickly and intuitively than expert users of digital cameras with limited experience with other digital products.

4.2 Remote Control Experiments

The present set of experiments - reported in detail by Blackler, Popovic and Mahar [2003b] - was designed to employ a software-based universal remote control to further test the thesis. These experiments will be the focus of the remainder of this paper. Because of the success of the previous work [Blackler, Popovic and Mahar 2003a], a similar method was employed.

Participants were selected from a heterogenous pool of volunteers from the QUT staff and their friends and family. None of the participants had encountered the remote control before and all were unpaid.

The prior research established that there is no significant correlation between the level of experience with a particular product type and the likelihood of a person being able to use the product intuitively. Grouping participants into expert, intermediate, novice and naïve with the test product seems to be less relevant when investigating intuitive use than other aspects of usability, because intuitive use involves applying knowledge from other contexts. [Blackler, Popovic and Mahar 2003a]. Therefore, Technology Familiarity score was the Independent Variable for the remote experiment. This was determined by the score on the Technology Familiarity questionnaire, which for this experiment was filled in by participants when they first volunteered to take part. This Technology Familiarity questionnaire included products with features similar to the remote rather than products with features similar to the camera. Using the TF scores, thirty people were assigned to three groups (high, medium and low levels of technology familiarity). Individual differences were controlled by selecting a cross section of people in terms of age, level of education and gender for each group.

Situational variables were minimised as much as possible. All experiments took place in an air-conditioned room with the same level of artificial light and the recording equipment, remote control, TV and VCR were set up and positioned in the same way each time. The participants were asked to complete three operations, each consisting of a number of tasks:

- Use the remote control to turn on the television and VCR and start playing the tape in the VCR
- Go to the start of the current recording (give name of program), play that scene for a few seconds and then stop the tape.
- Reset the clock on the VCR to 1724

Participants were asked to try and work the operations out for themselves before asking for the manual because using the manual masks use of relevant past experience.

5. Apparatus and Measures

The Marantz RC5000i universal touch-screen remote control (Figure 2) was programmed to control a Panasonic NV SD 220 VCR and NEC Chromovision TV. The operations were designed to investigate seventeen of the features on the default interface of the product, some of which are common to many digital devices, and others of which are found on most audiovisual equipment.

Figure 2. The remote control in use

5.1 Variables, Methods and Measurement Tools

Variables measured through this experiment and the methods and tools used are shown in Table 1.

Table 1. Variables, Methods and Measurement Tools

Dependant Variables	Methods and Measurement Tools
Time to complete operations	Observation using Observer software
Correct, inappropriate, incorrect and attempted uses of remote control features	Observation using Observer software
Conscious reasoning apparent during each use	Observation using Observer software Concurrent protocol
Number of first or only uses of features per participant that were intuitive	Observation using Observer software Concurrent protocol
Percentage of uses of each feature that were intuitive	Observation using Observer software Concurrent protocol
Familiarity of each feature	Structured follow up interview
Assistance received	Observation using Observer software

Two digital video cameras were used to record the activity, as per Vermeeren [1999] and Blackler, Popovic and Mahar [2003a]. One was trained on the participants' hands as they operated the remote, and the other recorded the whole scene. These pictures were digitally mixed to produce one MPEG file that showed both scenes. Following completion of the operations, participants were interviewed. The interview followed a structured format so that the results from each participant interview could be easily compared with the others [Sommer and Sommer 1997]. During the interview, participants were asked to explain which features of the product were familiar to them from other products they had used in any other situations. Each participant gave a familiarity rating (1-6) for each feature.

5.2 Coding Data

The Observer software was used to code the video and protocol, and produce quantitative data. The audio-visual data were coded as shown in Table 2. Every feature use for all participants was coded. All data were checked through twice to pick up any mistakes in coding, and the experimenter erred on the side of caution – an intuitive use was only coded where there was no doubt about it.

Table 2. Behavioural classes and modifiers used in coding

Behavioural class	Modifiers	Type
Feature uses	Correctness of use	Event
	Level of conscious reasoning	
Assistance received from manual	TV manual	State
	VCR manual	
	Remote manual	
Assistance received from Experimenter	Hint	State
	Detailed instruction	
Time on task	None	State

5.2.1 Correctness Modifier

A correct use was taken to be one that was correct for the feature and also correct for the task or subtask at the moment of use. A correct but inappropriate use was one that was correct for the feature but not for the task or subtask. In other words, users knew what they wanted to do and used the right feature to do it, but it was the wrong thing to do at that moment. An incorrect use was wrong for both the feature and the task, and an unsuccessful attempt was usually due to the participant not pressing the buttons on the touchscreen correctly, and failing to activate the feature.

When calculating the statistics relating to intuitive uses and intuitive first uses, only correct or correct but inappropriate uses were counted, because incorrect intuitive uses (of which there were very few) do not contribute to the successful operation of the product. Correct but inappropriate uses are relevant as this experiment was focusing on the features of the remote and these uses were correct uses of the particular features although they were not correct for the relevant task or subtask.

5.2.2 Conscious Reasoning Modifier

Conscious reasoning coding ranged from intuitive (fast decision with no evident reasoning), through quick comment (enough reasoning to verbalise a couple of words) and trial and error (random playing with buttons or exploratory behaviour), to reasoning (thorough reasoning evident), getting help (relevant past experience masked) and finally mistake (feature used by accident).

Table 3 shows examples of a reasoning use and an intuitive use of the four-way feature, which is used to manipulate a marker through a list on the TV screen in order to select the clock set function for the VCR. Both were correct uses on the first encounter of this function.

Table 3. Reasoning and intuitive uses of a feature

Reasoning use	I'll just experiment …I'm not sure. This changes the screen so I'll change...this is an arrow up so I'll change…ahh…demonstration…ah…language..clock set. I've reached the dot by clock set so that's the point of that dot there. OK, so it looks as though I'm getting there.
Intuitive use	Aha! OK here we go and I want to go to clock set. OK.

These examples show quite clearly how, although both participants were completing the same action, the level of reasoning is different for each. The second participant (the intuitive use) did not verbalise the steps required to get the marker to the clock set position so in this case the protocol lacks the detail of the reasoning process. The intuitive process is obvious from the protocol combined with the observation. Level of conscious reasoning was a modifier for the behavioural class "features" and was used to determine the number of intuitive uses and intuitive first uses. Numbers of intuitive uses and intuitive first uses were then used to calculate the results.

5.2.3 Results of Coding

The result of the coding exercise is a time-event log, showing all the behaviours and their appropriate modifiers in chronological order with a time stamp for events and a duration for states. Table 4 shows an example of part of a time-event log.

Table 4. Time-event log

Time (secs)	Behavioral Class	Behavior	Duration
3.08	Time on task	start 1	98.44
18.60	features	on, correct use, trial and error	0.00
23.48	features	navigation, correct use, intuitive	0.00
30.32	features	tv on/off, correct use, intuitive	0.00
41.00	features	volume/channel, correct use, intuitive	0.00
48.96	features	windows, inappropriate use, intuitive	0.00
56.88	features	navigation, correct use, quick comment	0.00

6. Results

These results have been reported in detail by Blackler, Popovic and Mahar [2003b] but a brief overview is provided here. The coded data were compared with the answers given during the interview to give two sets of results – those concerned with the features of the remote and their performance, and those concerned with the participants and their performance.

A strong, negative correlation was found between time to complete the operations and the technology familiarity score, $r(28) = -.658$, $p<.0001$ (all correlations are Pearson's product moment correlation coefficient). There was a strong, positive correlation between the number of intuitive first uses per participant and the technology familiarity score, $r(28) = .721$, $p<.0001$. And a strong, negative correlation between the number of first intuitive uses per participant and the time on the tasks, $r(28) = -.717$, $p<.0001$. Therefore, participants who had a higher level of technology familiarity were able to use more of the features intuitively first time and were quicker at doing the tasks. There was a strong positive correlation between time to complete operations and number of times help was used, $r(28) = .857$, $p< .0001$, and a significant negative correlation between TF score and number of times help was used, $r(28) = -.530$, $p< .003$. So, those with a lower Technology Familiarity score, as well as taking more time and using less of the features intuitively, also required more assistance.

The mean familiarity of the features correlated strongly and positively with the percentage of intuitive uses of the features, $r(15) = .698$, $p<.002$. The mean familiarity of the features correlated strongly and positively with the percentage of intuitive first uses of the features, $r(15) = .800$, $p<.000$.

7. Discussion

The correlations reported here between time, technology familiarity score and intuitive uses of the features support the findings of Blackler, Popovic and Mahar [2003a]. People seem to use their previous experience with similar features in order to use new features intuitively. These results also suggest that the decision to use technology familiarity score as the independent variable to group participants rather than level of expertise was the right one. The data on intuitive first uses are particularly important as they strongly suggest that people are able to use a feature intuitively the first time they encounter it if they are already familiar with a similar feature.

Features that were less familiar to users had less intuitive uses, more mistaken uses, more unsuccessful attempts and more incorrect uses. So using familiar features in new products should be the key to intuitive interaction. The detailed data on the features obtained from this experiment have enabled the re-design of the remote control in a systematic way that is aimed at increasing the intuitiveness of the product. The new design is being tested to determine how intuitive it is to use.

The methods used were very successful in getting the detail out of a lot of complex data and extracting the results. Although the coding is a time-consuming process, it is much faster than doing it manually and generates results in seconds. These authors believe that there would not at present be any other way in which they could extract this sort of detail from the rich audiovisual data and get such accurate and timely results from it. The fact that the results from experiments one [Blackler, Popovic and Mahar 2003a] and two [Blackler, Popovic and Mahar 2003b] largely agree suggests that the method has reliably pulled out the facts from all the raw data. A similar method has been used for experiment three to test the re-designed remote.

8. Conclusion

This research has the potential to help make products and systems intuitive to use. As well as the re-design and test process, future work could also include providing a framework for designers to help them apply intuitive use to their products as part of the design process. In order to investigate complex and unusual issues like this one, detailed and robust methods which can be tailored to suit each situation are needed. The issue of intuitive interaction involves human participants, an unconscious cognitive process and a very detailed focus on individual features. The method used here has twice proved to be successful in investigating this issue and gaining reliable results. The software particularly contributes to this, as it allows organisation and thorough analysis of very rich and complex raw data. Verbal protocol coupled with observation through this sort of software can be a successful method for studying complex and multi-faceted usability-based issues.

References

Agor, W. H., "The logic of intuitive decision making : a research-based approach for top management", Quorum Books, New York, 1986

Bastick, T., "Intuition: how we think and act", John Wiley and Sons, Chichester, UK, 1982

Blackler, A., Popovic, V. and Mahar, D., "The Nature of Intuitive Use of Products: An Experimental Approach", Design Studies, 24, No. 6, 2003a, 491-506.

Blackler, A., Popovic, V. and Mahar, D., "Designing for Intuitive Use of Products. An Investigation", 6th ADC, Tsukuba, Japan, 2003b

Bowers, K. S., Regehr, G., Balthazard, C. and Parker, K., "Intuition in the Context of Discovery", Cognitive Psychology, 22, 72-110.

Dreyfus, H. L., Dreyfus, S. E. and Athanasiou, T., "Mind over machine: the power of human intuition and expertise in the era of the computer", Free Press, New York, 1986

Ericsson, K. A. and Simon, H. A., "Protocol Analysis. Verbal Reports as Data", MIT Press, Cambridge, MA, 1984

Fischbein, E., "Intuition in Science and Mathematics", Reidel, Dordrecht, Holland, 1987

Jorgensen, A. H., "Thinking-aloud in User Interface Design: A Method Promoting Cognitive Ergonomics", Ergonomics, 33, No. 4, 1990, 501-507.

King, L. and Clark, J. M., "Intuition and the development of expertise in surgical ward and intensive care nurses", Journal of Advanced Nursing, 37, No. 4, 2002, 322-329.

Klein, G., "Sources of Power How People Make Decisions", Cambridge, MA, MIT Press, 1998

Sommer, B. and Sommer, R., "A Practical Guide to Behavioural Research. Tools and Techniques", Oxford University Press, New York, 1997

Wickens, C. D., Gordon, S. E. and Liu, Y., "An Introduction to Human Factors Engineering", Addison-Wesley Educational Publishers Inc, New York, 1998

Alethea Blackler, PhD Candidate and Part Time Lecturer
Queensland University of Technology
GPO Box 2434, Brisbane, Queensland, 4001, Australia
Telephone: +61 (0)7 3864 4334, Telefax: +61 (0)7 3864 1528
E.mail: a.blackler@qut.edu.au