Manuscript accepted for publication in International Journal of Human Computer Studies

How do we Program the Home?

Gender, Attention Investment, and the Psychology of Programming at Home

Alan F. Blackwell, Jennifer A. Rode[1] and Eleanor F. Toye

University of Cambridge Computer Laboratory

Corresponding author:

Alan Blackwell
University of Cambridge Computer Laboratory
William Gates Building, 15 J J Thomson Avenue
Cambridge CB3 0FD, United Kingdom

E-mail: Alan.Blackwell@cl.cam.ac.uk

Telephone: +44 1223 334418, Facsimile: +44 1223 334678

Abstract

We report a series of studies investigating the choices that users make between direct manipulation and abstract programming strategies when operating domestic appliances. We characterise these strategic choices in terms of the Attention Investment model of abstraction use. We then describe an experiment that investigates the estimation biases influencing the individual parameters of that model. These biases are linked to gender in a way that explains some gender differences in discretionary appliance use. Finally, we suggest design strategies that might compensate for those gender-linked estimation biases, and therefore make programmable features of future homes more accessible to a wider range of users.

Keywords:

End user programming, Gender HCI, Home automation, Domestic technology, Appliance design, VCR, Attention Investment

[1] Present address: University College London

Introduction

There has been a great deal of research investigating the digitally augmented "home of the future". This research has included the construction of many demonstration facilities showcasing new technologies and appliances, generous funding from manufacturers, and a few experimental observations of short-term residents as they learn to control such advanced domestic technologies. This has been complemented by important social and economic research understanding the context and consequences of technology deployment in the home (Harper 2003). The engagement of traditional computer science with this market opportunity has typically involved the transfer and adaptation of office technologies (such as networking and GUIs) to the slightly different application domains of home media consumption and social communication. A more recent trend has been the application of machine intelligence techniques to monitor and predict the behaviour of residents, thereby offering either predictive behaviour (for example smart alarm clocks (Isbell et al. 2004) and light switches, (Brumitt and Cadiz 2001)), or remote monitoring and surveillance (for example the remote monitoring of the elderly and disabled by family members and healthcare professionals (Consolvo et al. 2004, Mynatt et al. 2000)).

In our view, much of this research has neglected a valuable opportunity and requirement, that it will be necessary for somebody to program and configure home technologies. The appliances in a networked home will need to communicate with each other, and even stand-alone appliances need programming to define their future behaviour (for example, cooking appliances, and media recording devices). We expect to see significant growth of research interest in this problem, from many perspectives. For example, Grinter et. al. (2005) investigate "digital housework" such as the collaborative demands of maintaining advanced network technologies within a home. This research provides a valuable focus on technical challenge, complementing ethnographic studies of more commonplace and unremarkable practices within the home. In the research we present here, however, we take a specific technical focus on the challenges likely to be faced by individuals, from the research perspective of end-user programming. The objective is to provide design advice for those groups developing programming techniques for context aware ubiquitous computing (e.g. Sohn & Dey 2003).

If residents in the home of the future are to have significant control over their technological environment, they must be given the power to specify home automation functions. Many stand-alone devices already allow users to make a choice between direct manipulation (in which the effect of user actions is immediately visible as feedback) and abstract notation (in which the user expresses requirements in some form of language, defining behaviours that will take place in the future). Where appliances interact (for example through networking), the complexity of such choices can only increase in future.

In this paper we aim to provide insight into the extent to which domestic appliances should include abstract programming functionality, and whether this should be realised through the design of new interaction devices (such as the AutoHAN Media Cubes (Blackwell & Hague 2001)) or whether standard personal computers are adequate, in which case conventional approaches to end-user programming might be sufficient. We build on our work with Attention Investment (Blackwell 2002, 2006) and our program of research in the home looking at programming and gender (Rode, Toye & Blackwell 2004, 2005) which we will introduce in the following sections.

Attention Investment model of Abstraction Use

The Attention Investment model of abstraction use is a systematic account of the decision process involved when a user programs an appliance to do something ahead of time or repeat a complex sequence of actions, rather than simply achieving the same end by directly pressing the buttons at the right time. Typical programming tasks of this kind include programming a VCR to record a broadcast movie, or programming a speed dial code into a telephone.

People have different motivations for programming. Some people, such as computer scientists, program appliances because it is fun, or because it is their habit to explore all the esoteric features of everything they buy, or perhaps because they are so familiar with the kinds of procedure that will be involved as to find them extremely easy. In contrast, other people are very reluctant to engage in programming tasks, even if it might save them a lot of time and effort compared to doing things by direct manipulation. The decision process is therefore a cost-benefit analysis, typical of Simon's (1956) rational choice models of human behaviour. Simon observed that people seldom take the time to find an optimal solution, instead "satisficing", thinking about the problem only long enough to find a solution that is satisfactory rather than optimal. In the same way, people might choose to carry out actions manually by direct manipulation, rather than creating a program to automate them, if this will require less mental work. In the Attention Investment model, the "utility function" that is being optimised is a function of cognitive effort, which might be informally described as a quantity of concentration, but which we describe simply as "attention".

Programming and direct manipulation represent two alternatives for completing a task, where the alternatives differ in terms of the time and attentional resources required. Programming involves a certain amount of concentrated attention to understand or form a suitable abstract specification of the required action, whereas direct manipulation usually involves a longer period of less effortful attention. The cost part of the Attention Investment equation is the amount of up-front time and attention required if the abstract alternative is chosen. The return part of the investment equation is the saving that this produces, by reducing the amount of time and attention that would otherwise be occupied in future direct manipulation. However this is not a guaranteed investment. If the program is incorrect, or the specification is faulty, then the anticipated savings might not be achieved, or might even result in further costs in future (for debugging or repair). Of course, this risk can be reduced by further up-front effort, analysing the situation, reading manuals, testing, and so on. But all such effort requires further investment of time and attention, and reduces the proportional "profit" that will result from having chosen an abstract strategy.

Most users do not spend a long time deciding between the likely benefits of an abstract programming strategy rather than a direct manipulation strategy. There are so many possible considerations involved that even thinking about the right strategy might take more time and effort than the task itself. The greatest payback for attention investment might be not to spend time on this decision at all, but simply to make the same choice you did last time you made a similar decision. This is described by cognitive scientists as a "bounded rationality" model, where the amount of time spent on the rational decision process is bounded by a limited amount of time available for reasoning. In terms of human behaviour, bounded rationality models account for the fact that people do not always want to spend a long time weighing up alternatives, but often act instead on the basis of heuristic shortcuts or previous biases (Russell & Wefald 1991, Gigerenzer & Selten 2001). In previous work we implemented a bounded rationality simulation of decisions between direct manipulation and abstract strategies that confirmed this as a plausible model of the Attention Investment theory (Staton 2005). If rational reliance on biases leads users to repeat previous choices, then programmers will tend to choose an abstract strategy, while non-programmers will tend to choose direct manipulation.

The Attention Investment model is related to other descriptions of end-user strategy such as Carroll and Rosson's Paradox of the Active User (1987) which describes the way that users are reluctant to suspend productive use of already-learned (but perhaps inefficient) methods, and tend not to engage in learning further skills, even though this might bring longer-term benefits. Attempts have been made to accommodate end-user programmers who may have these biases, for example by supporting "just-in-time" programming (Potter 1993), a design approach which encourages users to make small steps toward programming, rather than demanding up-front investment. The general phenomenon has been described in more qualitative terms as the "gentle slope" of programming and tailoring (MacLean et al. 1990, Pane & Myers 2006), whereby the programming tools make it easy to do simple things, with only gradual increases in difficulty as the programming task becomes more challenging. This is considered a desirable property of an end-user programming environment, which should be easy to use for simple things, but also sufficiently powerful to support complex behaviour. The specification of complex behaviour will obviously be more complex, and thus more difficult for users. However, system designers should aim for a gentle slope of rising difficulty as the complexity of the user's goals increases. A sudden increase in difficulty, especially if not motivated by the complexity of the goals, is likely to cause users to abandon the system at that point. These phenomena can be described more quantitatively in terms of the Attention Investment model, as we do in the experimental study described later in this paper.

In the present study, we aim to identify design approaches for consumer electronics in a way that can help users in the home decide between abstract and direct manipulation strategies for the control of home appliances. We are particularly interested in the case where users' previous biases might deter them from choosing the most appropriate strategy for a particular task. One design approach to mitigating this problem, that was previously motivated by the Attention Investment model, is the "surprise, explain, reward" technique developed by Wilson et al. (2003) for use in spreadsheet design. The "surprise" component is intended to distract users from bias toward a habitual strategy, the "explain" provides enough information for the user to reassess the attention investment factors, and the "reward" makes clear to the user what the payback has been for that investment choice, in a way that will influence future choices of strategy.

The primary goal of our research is therefore to evaluate whether the Attention Investment model is appropriate to the home domain, and see whether it is descriptive and predictive in this context. If so, a secondary goal is to provide design advice, founded in empirically validated theory, that is relevant to consumer electronic design.

Prior Work: Programmable Appliances in the Home

Research by Blackwell et al. shows that consumer electronics interfaces for the home abound with abstract functions in their user interfaces (Blackwell, Hewson & Green 2003, Blackwell 2004). That study of consumer electronics interfaces in the UK's largest electronic retail chain found, for example, that audio mini systems, in addition to the various media controls, had an average of 10.7 further controls that were used for manipulating abstract representations. This might include functions such as programming an alarm to go off at a particular time and play a particular radio stations. That prior research suggests that abstraction is a significant trend in consumer electronics, and that users' decision making practices in whether to engage in direct manipulation or abstraction are increasingly relevant.

"Fuzzy Felt Ethnography": Contextual study of gender and abstraction

To complement Blackwell's study of commercial consumer electronics interfaces we started our research with contextual studies of home programming activities, carried out in people's homes. It was essential that these studies were conducted in a manner sensitive to the contextual influence of strategic bias. The first of these, previously reported in Rode, Toye and Blackwell (2004), took particular care to dissociate the research from the technical environment of programming and computer science. As the title of that paper indicates, we carried out a field study in which potentially programmable appliances were represented by icons made from felt. These were used to discuss technical behaviours in a non-threatening way, sitting with a family in their own home after eating a takeaway meal with them. Programming was not mentioned explicitly by the researchers, but informants were instead invited to consider ways in which their own household appliances behaved in ways that could not be accounted for by the conventional usability principles of direct manipulation. This included any means of programming easily "repeated" behaviour (e.g. macros or shortcut keys), and means of defining automated behaviour "ahead of time" (e.g. timers or event triggers).

We interviewed the members of nine households, screened to exclude households that contained professional computer programmers and computing researchers, since we expected that programmers would tend already to have biases in favour of programming strategies. Participants included younger and older couples, singles, and families with children. The adults were aged 29-60, with a mean age of 40 years. The numbers of electronic appliances counted in a household ranged from 22 to 55 (mean 34.2). There were some duplicates found, but the mean number of different appliance types we found in a household was 29.4. A surprisingly large proportion of these could be programmed in some way. We found between 4 and 10 different *kinds*[2] of appliance in each household

[2] A single generic kind of appliance, as reported in our analysis, might be represented by several individual appliances in any given household.

that could be programmed ahead of time (mean 5.9) and between 1 and 5 different *kinds* of appliance that could be programmed to repeat behaviour (mean 3.2).

Not all appliances capable of being programmed were actually used in this way. Although a home contained 5.9 kinds of appliance, on average, that could potentially be programmed ahead of time, the average number that actually were programmed across the homes we surveyed was only 3.9. Although 3.2 kinds of appliance, on average, could be programmed to repeat behaviour, the average number that actually were programmed was only 2.0. We asked participants to assess the relative difficulty of programming these various appliances in their home, and the frequency with which they actually did so. We did not find any significant difference between the two kinds of programmable functionality, either in terms of relative frequency of use or perceived difficulty, despite the anecdotal difficulty of particular tasks (programming a VCR, for example).

We also wanted to understand the relationship between gender roles and domestic programming, as preliminary research had indicated that attitudes to home technology differ significantly with gender. For example, Livingstone (1992) has outlined a number of key differences in how men and women discuss domestic technologies. In her personal construct interviews, women talked more concretely about the significance of domestic technology in their lives. For example, women used the word 'control' to mean minimizing domestic chaos, while for men, 'control' was a factor that allowed them to express personal expertise. Men tended to emphasize that technological objects were 'purely functional,' and discussed them in terms of their inherent properties and features. However women are more concerned with the utility of objects – whether they make tasks easier (Livingstone 1992, p. 120). Although Livingstone's interviews did not address programmable technologies, we expect that the attitudinal differences she observed would influence the Attention Investment model. In particular, we would expect that men are more likely to invest in exploring and programming 'risky' new features, while women would require more certainty that there would be a useful labour-saving 'profit'.

Subsequent work by (Beckwith et al. 2006) has demonstrated that gender-linked performance in programming can be attributed to the fact that gender is correlated with self-efficacy, the quality defined by Bandura (1986) as a person's prior belief in his or her own ability to accomplish a particular task. According to the Attention Investment model, this factor would be expected to have direct consequences on programming behaviour, because low self-efficacy would lead to unrealistically high estimates both of attention cost, and of the risk of failure. In our own study we did not find any overall difference between males and females in willingness to program appliances, or in perceived difficulty of programming. We did, however, see a difference in the types of appliance that were programmed by males and females.

This difference between types of appliance can be characterised in terms of Livingstone's observations about gender differences in attitudes to utility and control. In our interviews, men reported that they found programming easier for equipment that was primarily recreational, such as CD players and car radios, whereas women were more comfortable with programming devices that permitted them domestic control: alarms, ovens, heaters, bread-makers, security systems etc. In terms of Attention Investment theory, we would expect that repeated experience of such programming operations would cause men and women to develop differences in their estimates of attentional effort and risk of failure when selecting between manual or programming strategies for entertainment or household management appliances respectively.

This follows gender roles where women still tend to be responsible for coordinating and doing the majority of domestic work, whereas men have more opportunity for leisure (Bureau of Labor Statistics, 2004; Antonides 1998; Maushart 2001; Berk 1985; Deem 1986). Consequently, we hypothesized end-user programming specialization in the home followed typical patterns of gender use. We noted, for example, that the appliances programmed by men are primarily those devices that are programmed to repeat, whereas those used for domestic control consist of primarily ahead-of-time devices. However the VCR proved to be an exception. Despite the fact that it was an ahead of time device, and one for which more women were engaged in its programming, nevertheless men felt it was significantly easier to use than women, perhaps because of its recreational media function.

Diary Study: The VCR as a problematic case

In order to investigate the interesting status of the VCR as mentioned above, and also to investigate the ways in which attention is invested in abstract rather than direct control in the home, we conducted a follow-on study of VCR programming in the home. By focusing on an existing appliance category that is already available in most homes we were able to study the generic applicability of the Attention Investment model, without confounding other factors that influence next-generation technology adoption. We conducted a two week study of VCR usage by eight families (Rode, Toye & Blackwell 2005). Each household member completed a diary of their VCR use during this period, and 45-minute interviews were conducted with each household at the start and end of the period, in order to explain and interpret those diaries. There were 16 adult participants in the 8 households, aged from 25 to 63, with a mean of 42.8 years. The households also included 15 children, 8 girls and 7 boys, ranging in age from 18 months to 17 years. As in our study of appliance ownership, all participants were screened during recruitment in order to ensure that they were non-programmers, and non-computer scientists.

Four of the households did not make any recordings during the study period. We were able to collect other data on patterns of VCR usage, but did not further investigate programming. Of the four households who did make recordings during the study period, 44 recordings were started manually, while 9 were programmed to start recording under timer control. This preference for direct manipulation provides some confirmation of the anecdotal belief that VCR programming is too hard, if users prefer to invest their attention in the need to be present at the time recording should be started. In the cases where recording was programmed in advance, we found that particular individuals in these households had assumed responsibility as the expert user of the VCR functions. This expertise resulted in a degree of skill-trading, or (in attention investment terms) trading of attentional effort, as a component of the domestic economy (for example in return for other household responsibilities). Programming was not, as popular stereotypes suggest, solely the responsibility of teenage boys (although we did observe this pattern), but was used as a mechanism for household control, for example when mothers recorded children’s television in order to manage the time of young children through time-shifted programme viewing.

As with our survey studies, we found that the configuration and customisation of home technologies, when distinguished from direct manipulation control, is implicated in rich patterns of behaviour within the household. The Attention Investment model of abstraction use describes personal choice between programming and direct manipulation strategies in cognitive terms, but does not take account of extrinsic motivations such as those arising from the social context in which these tasks are carried out, and by which they might be motivated.

This study also left us with a significant question related to gender. In the "Fuzzy Felt Ethnography" we learned that men considered VCRs significantly easier to use. This diary study confirmed our concern that women might be likely to overestimate expected effort and risk of failure when using VCRs. We wanted to better understand this disparity, test how it was related to the choice between abstraction and direct manipulation in the Attention Investment model, and see how it changed as users became more experienced with a device.

Attention Investment for a home media appliance

In this experiment we selected a specific programmable home appliance, in order to investigate the parameters of the Attention Investment model as they apply to domestic appliances, but in a more controlled environment than the homes in which we conducted our field work. We were particularly concerned with the parameters that influence the Attention Investment model, in order to investigate whether the model is sufficiently complete to account for the cognitive factors in media appliance programming.

We therefore wished to explore, for a population of non-programmers, how they would estimate their own performance and likelihood of success when faced with the programmable functions of a new appliance. We asked them to attempt a basic programming task, and assess the results of that attempt. This experiment was conducted in a context designed to be as much as possible like somebody purchasing and using a new appliance in their own home.

The appliance that we used was chosen to be typical of newly acquired home appliances, so that it was a new programming experience, and yet was selected to be in the same general class as the VCR. We therefore chose to use a relatively familiar, but nevertheless novel, example of current generation retail products. This was a DVD player and recorder that included typical features to support programmed recording of broadcast TV. The model used did provide on-screen programming instructions, which are typical of similar products. The market for this type of device is changing rapidly, and they are rapidly being replaced by Personal Video Recorders such as TiVo or (in the UK) FreeView recorders. Different products may well be more or less well-designed than the one we used in our experiment, but this is not particularly important for our research concerns, which are to do with the relationship between perceived and actual performance when estimating the difficulty of programming tasks.

We asked our participants to work out how to use the DVD recorder and television provided in our lab, to record a programme from the TV schedule – in one experimental condition by direct manipulation and in a second condition via an abstract specification (i.e. setting up the DVD recorder to record a TV programme at a later time). Our aim was to collect data concerning the participants' actual performance on these two tasks, as well data about their beliefs and previous experience in relation to domestic technology in general, and this type of technology in particular. We would then be in a position to compare the participants' impressions of the DVD recorder, the two tasks, and their own ability and previous experience of direct manipulation and abstraction, with their actual performances when using the DVD recorder.

Method

Participants

Twelve men and twelve women took part in this study. As in previous studies, we recruited participants with no professional experience of computer programming. Participants were recruited through word of mouth, and through posters displayed around the town and university. They were paid for participation with a 20 pound book token (gift certificate for books).

The mean age of the participants was 30 years and 5 months (range 23 years to 47 years 8 months, standard deviation 6 years 11 months). The ages of the men and women were matched pair-wise, so that for each pair, there was no more than a four-year age difference. The mean age of the women in the study was 30 years 8 months (range 24 years 3 months to 43 years 6 months, standard deviation 6 years 7 months) and the mean age of the men was 30 years 2 months (range 23 years to 47 years 8 months, standard deviation 7 years 7 months). No participants reported any health difficulties that might affect their ability to use a DVD recorder. One woman participant was not a native English speaker, but had been living in the UK for the past 10 years.

None of the participants were computer science students or researchers, or had ever been involved in designing TV recording technology. Four participants (all male) had previously made recordings to a DVD disk, but in all four cases, the DVD drive was embedded in a PC, and was being used for recording computer data. None of the participants had previously used a DVD recorder to record television programmes.

Apparatus

Figure 1. Usability lab with domestic furnishings: a) TV and DVD recorder on TV stand, with participant sitting in armchair to use remote controls; b) one-way glass window separating observation area (behind and to the left of the participant's armchair, from which this photograph is taken).

The study was conducted in a small usability lab at the University of Cambridge Computer Laboratory. The usability lab was configured to provide a main 'living room' area with domestic furnishings and a small observation area, from which the main area can be seen through a one-way mirror. In the main area, a TV and DVD recorder were placed on a TV stand. Both the TV and DVD-recorder were made by the same manufacturer, Panasonic (model numbers TX-14B4T and DMR-E60EB respectively). The front panels and remote control layout for both are illustrated in Figure 2. At the time of the experiment, DVD recorders were relatively exotic appliances, and the recorder was significantly more expensive than the TV. However, we expected that prices of DVD recorders would soon fall (this has indeed happened), and that the two models could be regarded as comparable "standard" consumer appliances, rather than specialist devices marketed to technical enthusiasts. They were purchased together from a department store (rather than a specialist electronics supplier), with advice from a sales assistant confirming our requirement that these were standard products that would regularly be sold to customers without any specific technical expertise.

a)
b)
c)
d)

Figure 2. Experimental appliances: a) Panasonic DVD recorder model DMR-E60EB, b) DVD recorder remote control, c) Panasonic television model TX-14B4T, d) TV remote control

The two devices were correctly connected to each other. Both were connected to the electricity supply and left in ‘standby’ mode. The set-up procedures for the TV and the DVD recorder had both been completed by the researchers, so the devices were ready for immediate use. For both the two recording tasks we set our participants, the simplest correct procedure was as follows (Note that these steps are listed for the benefit of the reader in understanding the nature of the two tasks – participants were not given these instructions):

1) Turn on TV using TV or DVD remote control;
2) Turn on DVD recorder, using either the buttons on the DVD-R front panel, or on the DVD-R remote control;
3) Switch the TV to AV mode, using the AV button on either the TV or the DVD remote control[3];

[3] In order to complete the first task, participants had to turn the two appliances on (from standby mode), and learn to use the remote control to control the on-screen menu. Doing so required changing the mode of the DVD player so that its video signal was output to the TV—this was done by pressing “TV/Video”. Our previous involvement in

4) Open the DVD drawer by pressing the "OPEN/CLOSE" button on the DVD-R front panel, insert the blank DVD disc correct side up and close the DVD drawer.

For the simple record task, having completed steps 1-4 above, it was necessary only to tune the TV to the desired channel and press the 'record' button, on either the DVD-R remote control or on the DVD-R front panel.
For the pre-record task, the simplest next steps were as follows:

1) Press the 'Functions' button on the DVD-R remote. This results in a menu being displayed on the TV screen;
2) Use the arrow keys on the DVD-R remote control to navigate the on-screen menu, select Timer Recording from the menu and press Enter;
3) The Timer Recording menu is then displayed. Enter the programme details requested on the TV screen (channel, programme name, start and finish time, date) using the navigation buttons on the DVD-R remote control, and press Enter;
4) The TV picture returns, and a message is briefly displayed centrally on the screen which says 'Now press Timer Record button';
5) In fact there is no 'Timer Record' button on the remote control, but there is a button labelled 'Timer'. Press the Timer button. The TV screen will then go blank until the start of the programme being recorded.

These step-by-step instructions were not given to the participants, who nonetheless did have access to the manuals for the TV and the DVD recorder. The manuals were placed on the stand next to the TV, along with the supplementary set-up manuals, remote controls for both devices, and a blank DVD disc in its case. In addition to the TV trolley with TV and DVD recorder, the usability lab contained a comfortable chair for the participant, several upright chairs, and a table. The room was furnished to look pleasant and comfortable, and as similar as possible to a real living room.

Procedure

Each participant was told that they would be asked to attempt two tasks using a typical DVD recorder – recording from TV straight away, and setting up the DVD recorder to record a TV programme at a later time. The order of these two tasks was counterbalanced across participants, so that 6 men and 6 women did the simple record task first, and 6 men and 6 women did the pre-record task first. Participants were told that for both tasks, we were primarily interested in how generally easy or difficult it was to use the equipment, and that we were not assessing the participants' own technical ability.

The DVD recorder, TV, remote controls, DVD disc and manuals were then pointed out and introduced. For the pre-record task, TV schedule details were provided and pointed

---

the consumer electronics industry suggested this would be problematic, and it was to mitigate the problem as much as possible that we purchased two products from the same manufacturer, ensuring interoperability between the on-screen menu systems. Although relatively trivial, this proved a setback for many participants at the outset of the task. It was usually solved by reference to the manual, but it did result in an immediate loss of confidence. This confidence loss was relevant to our hypotheses, but is an intrinsic problem with modern consumer electronics.

out along with the manuals. Participants were told that the equipment was already set up ready for use and switched on at the mains. They were then briefly interviewed before starting their first task. The script for this interview is included in Appendix A.

We also asked participants to predict how long they thought it would take them to complete the task (by selecting one of a number of time-range choices), rate how confident they were about completing the task successfully, and rate how difficult they thought the task would be. Note that in order not to bias participants with the expectation that the task was going to be difficult, the questionnaire asked for an assessment of how easy or difficult the interface was to use, with a 9 point scale on which 9 was very easy and 1 very difficult. This avoids a common experimental bias toward the easy end of a scale However in the following analysis, we refer to "difficulty" (= 10 – response), in order to facilitate discussion of relationship between difficulty, effort and elapsed time.

Participants were advised that if they 'got stuck' on the task, they should make a reasonable effort to solve the problem, but that if they were sure they could not make any further progress, they could give up. As described in the results section, several participants did choose to give up. They were also told they could take a break during the task if they needed to for any reason, with the clock stopped during that time, although no participant did in fact make this request.

The researcher retreated to the observation area, and then gave the participant a verbal signal to start the task, simultaneously starting the clock. While the clock was running, the researcher made notes on the participant's actions, in order to be able to refer to these in the post-task interview. When the participant stated that they believed they had completed the task, or that they had decided to give up, the researcher stopped the clock and noted the elapsed time. It should be noted that it was possible for the participant to believe they had completed the task when they had not in fact done so. In all cases, the final states of the DVD recorder and the TV were noted.

Participants were interviewed after the first task, asking how long they believed it had taken them to complete (or attempt) the task, to rate how easy or difficult they found the task, and how confident they were that they were successful in completing the task. They were then asked an open-ended question about why they believed they had been successful or unsuccessful, or why they were unsure, and their answers were noted. Finally, they were asked an open-ended question as to what factors made the task easier or more difficult for them.

The participant was then asked to wait outside briefly, while the researcher checked whether they had completed their task successfully and re-set and checked the equipment for the second task. If the participant believed they were successful in recording when they were unsuccessful the experimenter informed the participant, and gave them an opportunity to comment on why this might have happened.

The procedure for the second task, including the interview questions and post-task equipment checking and re-setting, was identical to the procedure for the first task.

Following the two tasks and the interview questions immediately related to the tasks themselves, a further interview was conducted. This was to gather some background information about each participant's experience with and opinions about TV recording technology and other technology generally, and to ask for their assessment of the design of the DVD recorder they had just used, relative to other technology more familiar to them. The interview included a mixture of open-ended questions to allow participants to comment freely, and multiple choice questions for which the responses could be easily counted and classified. The interviewer script is included in Appendix B.

Results

The presentation of results is organised in accordance with our overall research questions, each of which is addressed in a separate sub-section below. To summarise these questions, we were interested in:

Performance measures: how did participants differ in their ability to carry out the simple-record and program-ahead tasks?

Perception of performance: regardless of the *actual* performance on the two tasks, how did participants *expect* to perform and *evaluate* their performance? Were their Attention Investment judgements incorrect, as a result of systematic biases in estimation?

Effects of gender on estimation: If there were systematic biases in perception and estimation, are these related to gender in a consistent manner?

Effect of gender on interpretation of experience: Did participants accurately modify their perceptions and estimates on the basis of experience? Did this modification compensate for any previous biases, especially those related to gender?

Finally, we also report a post-hoc test, in which we consider whether subsequent research on gender-linked experience with spreadsheets might have provided positive opportunities for the recording domain that were not anticipated in our own research questions. Findings from our qualitative analysis of questionnaire responses are also reported at points where they have a bearing on other experimental findings.

The same statistical techniques are used in all following sections. Note that whereas the actual task performance is a continuous measure of time taken, the time estimates made by participants were obtained using a multiple choice question, as shown in Appendix A. The results were not normally distributed, as is often found with participant responses using a multi-choice scale. Non-parametric statistical tests were therefore used, and form the basis for statistics quoted in the following sections. Non-parametric tests were also used for estimates of difficulty, following standard practice for subjective data collected using Likert scales. For all within-subjects tests (e.g. comparing the simple-record and program-ahead task), the Wilcoxon signed ranks test was used to make non-parametric paired sample comparisons between the two values for each participant. For all between-subjects tests (e.g. gender differences), the Mann Whitney U test was used to make independent sample comparisons.

Performance Measures

The experiment included two task performance measures. The first of these was successful achievement of the specified task (i.e. whether the recording parameters had been set in a way that would record the required TV show). The second was the time taken to complete the task (that is, task completion as judged by the participant, whether the record settings were correct or not). To review the experimental design, there were three independent variables: one within-subjects and two between-subjects. The within-subjects variable was type of task, either simple-record or program-ahead. The first between-subjects variable was the order in which these two tasks were carried out: either the simple-record task first, or the program-ahead task first. The second between-subjects variable was the gender of the participant. We carried out a mixed-measures analysis of variance (ANOVA) to evaluate our hypotheses regarding the effect of these variables on task completion time.

We expected that the programming-ahead task would be more difficult for all participants, regardless of which experimental condition they were in (program-ahead first, or simple-record first). This was confirmed by our results. Average time to complete the program-ahead task was 16:42 minutes across all participants, and for the simple-record task was 6:42 minutes. ANOVA showed that this main effect was statistically significant ($F(24,2)=32.8$, $p<.001$). There was no main effect of order: average total time to complete both the tasks did not differ significantly whether the simple-record or the program-ahead task came first.

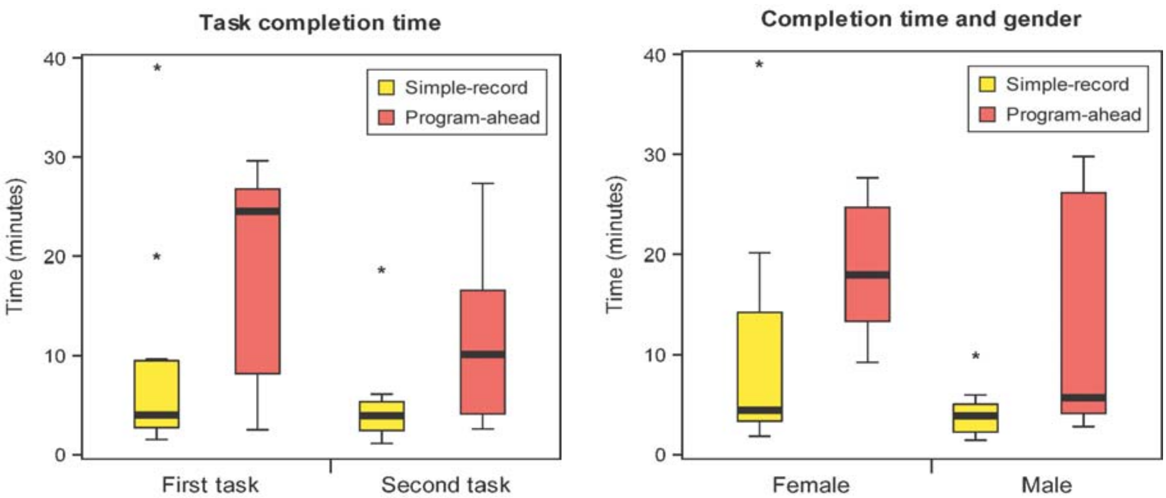

Figure 3. Overall task completion times: a) average time for the two tasks, showing the effect of order within the experiment in which each task was encountered; b) average times taken by male and female participants. One extreme value (female taking over one hour for first task program-ahead) has not been shown, in order to improve visual clarity on the vertical axis, but all data was included in statistical analysis.

We did expect that participants' performance would improve after practice with the DVD recorder. In addition, for those participants who were delayed by working out how to display the onscreen menu, the first task to be done was further prolonged by that delay. These expected and observed causes of delay were confirmed by the ANOVA, in the form of an interaction between task and order condition: each task took longer if it came first, and was completed faster if the participant had gained some practice in use of the appliances and the onscreen menu by doing the other task first (F(24,2)=7.35, p<.05). These effects are shown in Figure 3a).

We expected an effect of gender on task performance. Female participants did take longer to complete both tasks, as shown in Figure 3b) (F(24,2)=7.6, p<.05). This time discrepancy might be explained by Beckwith's theory that females are more likely to engage in reflection during programming tasks (Beckwith et al. 2006). We also expected that female participants might take even longer to complete the program-ahead task, especially in the condition where that task was performed first. However this hypothesis was not confirmed, with the ANOVA showing no significant interaction between gender and task, or between gender and order.

Simple record	**Total**	Males	Females	First task	Second task
Was User Manual Consulted?					
Yes	**19**	7	11	9	10
No	**5**	5	1	3	2
Was Task Completed?					
Yes	**23**	12	11	11	12
Gave up	**1**	0	1	1	0
Was stopped	**0**	0	0	0	0
Was Task Successful?					
Yes	**18**	9	9	9	9
Partial	**6**	3	3	3	3
Fail	**0**	0	0	0	0

Program-ahead	**Total**	Males	Females	First task	Second task
Was User Manual Consulted?					
Yes	**20**	9	11	9	11
No	**4**	3	1	3	1
Was Task Completed?					
Yes	**19**	10	9	8	11
Gave up	**4**	2	2	3	1
Was stopped	**1**	0	1	1	0
Was Task Successful?					
Yes	**18**	10	8	7	11
Partial	**0**	0	0	0	0
Fail	**6**	2	4	5	1

Table 1. Summary of User Manual consultation, Task Completion, and Task Success for Simple and Pre-Record Tasks.

With regard to success in completing the task, male and female participants were equally successful in completing the simple-record task, as seen in Table 1. Male participants were slightly more likely to succeed in the program-ahead task, but a Chi-square test does not find this effect to be statistically significant.

Results: Perception of Performance

Our main concern arising from the Attention Investment theory is not with *actual* task completion times or success, but with *perceived* task completion times and likelihood of success. Strategy choices arising from Attention Investment decisions are made on the basis of estimated costs and risks, because actual costs and risks are not available until after the task is complete. In our experiment, participants did not have any choice of strategy (we told them which order they should perform the tasks in). However, we chose to assess the likely basis by which they would have made a strategic choice by asking them to estimate, before carrying out a task, how long they thought it would take, estimate the effort involved, and whether they thought they were likely to succeed. We also assessed the likely basis of future choices by asking them to estimate, after each task was completed, how long it had taken them, how difficult it had been, and whether they thought they had succeeded. Any systematic discrepancies between actual performance and perceived performance would form the basis of future estimation biases in attention investment decisions.

In order to validate these estimation measures, we compared them to actual performance by the participants.

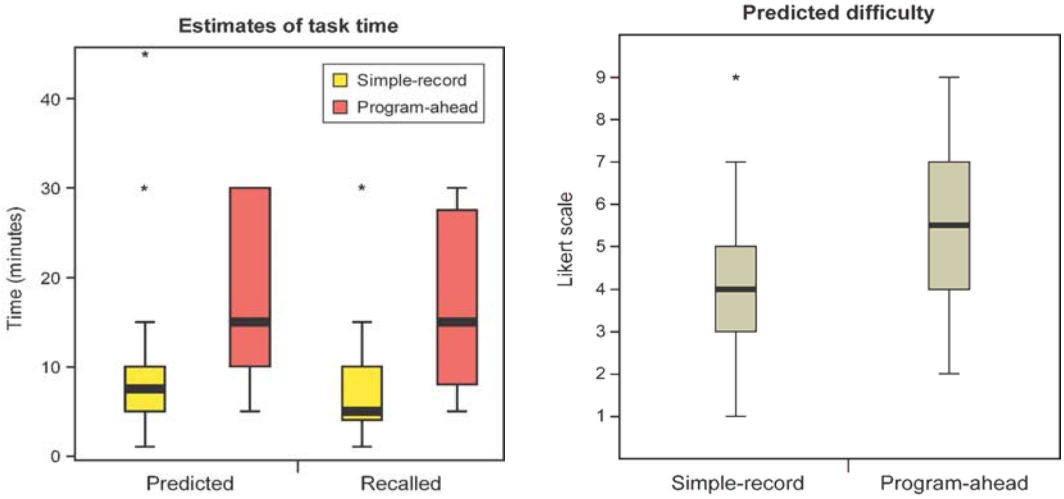

Figure 4. Participant's estimates of task duration and difficulty: a) estimated duration of the two tasks, made before and after the task was performed; b) predicted task difficulty. Two extreme time estimates (over one hour for prediction and recall of program-ahead time) has not been shown, in order to improve visual clarity on the vertical axis, but all data was included in statistical analysis.

As noted in the previous section, the actual time taken to complete the program-ahead task was longer than the simple-record task. Estimates made by participants indicated that they expected this to be the case. As shown in Figure 4a), the advance estimate of time to complete the program ahead task was higher than for the simple record task (Wilcoxon signed ranks, $p<.001$). We also found that the estimate of time taken after carrying out the task was higher for program-ahead than for the simple-record task (Wilcoxon signed ranks, $p<.01$). As shown in Figure 4b), we found that predicted effort for the program-ahead task was that it would be more difficult than the simple-record task (Wilcoxon signed ranks, $p<.001$). Assessment of effort after the task had been completed did not show a significant difference between the two tasks, largely because of high variability within each task. Some of this variability is accounted for by gender differences, as discussed in the next section.

These results show that participants are able to make reasonable estimates of the factors that contribute to the Attention Investment model, in that their assessment of the relative difficulty and time for different tasks did correspond approximately to the actual time that it took them to complete those tasks. The results also show that, as predicted by the Attention Investment model, participants expected program-ahead operations to require more time and effort than other functions of the same device.

Results: Effects of Gender on Estimation

As in our previous work, we next analysed the extent to which our experimental participants behaved in ways that are correlated with gender. Qualitative analysis of questionnaire data allowed us to confirm that participants in this experiment are representative of the same gender characteristics that we have observed in previous research. As in previous research (Rode et. al. 2004), we found that both men and women were more likely to describe experience with appliances identified with their own gender. Where female participants did describe experience with "masculine" appliances, these were most likely to be PCs. Where male participants did describe experience with "feminine" appliances, these were most likely to be washing machines and microwaves. Overall, fewer women were likely to volunteer evidence or make claims of technical competence.

Previous work by Beckwith et al. (2006)[4] has shown that the decision factors in the Attention Investment model (i.e. attentional effort and risk) can be estimated differently by different users, in a manner that is apparently biased with respect to gender. The Attention Investment model predicts that this will have consequences for strategy choice between abstract and direct manipulation strategies, such that if females over-estimate effort, and under-estimate likelihood of success, they will therefore be less likely to choose abstract strategies. Beckwith also found that males are likely to under-estimate effort and over-estimate likelihood of success for programming tasks, and as a result are *more* likely to choose abstract strategies. We are of course interested to test whether these same effects are seen in the use of domestic programming technologies, just as in the more business-oriented domain of spreadsheet programming that was investigated by Beckwith.

[4] The publication by Beckwith et al. is previous to this publication, although her study was in fact carried out after this one.

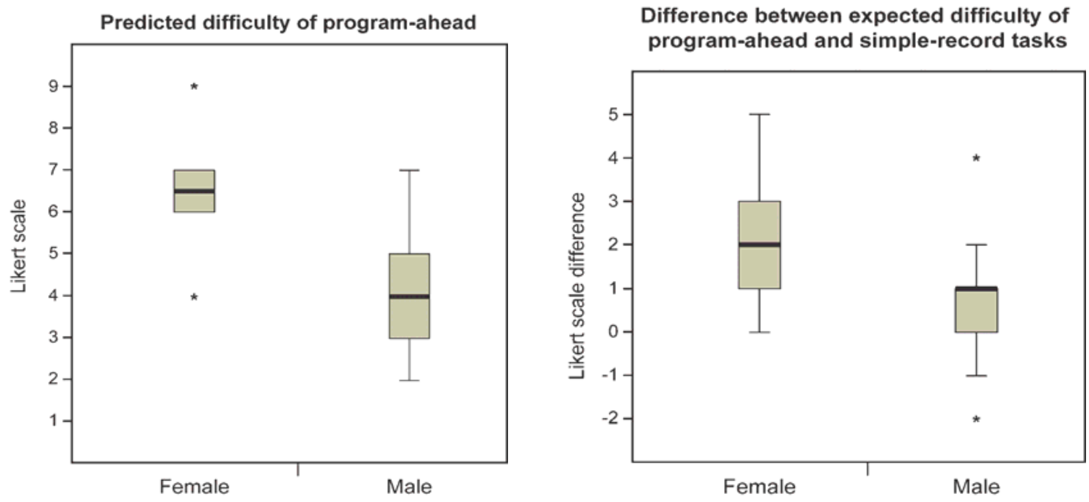

Figure 5. Participant's estimates of relative difficulty of the program-ahead task: a) female participants estimate that program-ahead will be more difficult than males; b) females perceive a larger difference in difficulty between simple-record and program-ahead.

We therefore compared the difference between the estimated effort ratings for the simple-record task and the program-ahead task, as shown in Figure 5. As noted above, all participants expected the program-ahead task to be more difficult than the simple-record task (Figure 4b), but we hypothesised that this difference might be more pronounced when estimated by females. This hypothesis was confirmed, as shown in Figure 5b) (Mann Whitney U, $p<.05$). We also found that males generally expected the program-ahead task to be easier overall, as shown in Figure 5a) (Mann Whitney U, $p<.01$). There was not a significant difference between males and females in expected difficulty of the simple record task.

We expected that there might be a difference between males and females in estimated likelihood that they would succeed in the programming task (i.e. the risk factor in the investment model). However, this hypothesis was not confirmed – there was no significant difference in the estimated likelihood of success. Qualitative analysis of questionnaire responses did show a gendered difference in how participants talked about their experience of completing the task. Females used less confident terms and descriptions than men. We also noted that females tended to discuss the device itself in terms of its utility and benefits, while males discussed it in terms of product features, an observation that is consistent with the findings of Livingstone (2002) discussed earlier.

Results: Effect of Gender on Interpretation of Experience

Our other research question relates to the effect of experience on future attention investment decisions. Successful completion of a task should build confidence, which we would expect to encourage users toward choosing more abstract strategies in future. We can measure this by comparing the estimates that participants made of the time that they had taken after a task was completed. We are particularly concerned that, even in the most favourable condition where females undertake the simple-record task, the evidence suggests that they still do not assess their achievement realistically.

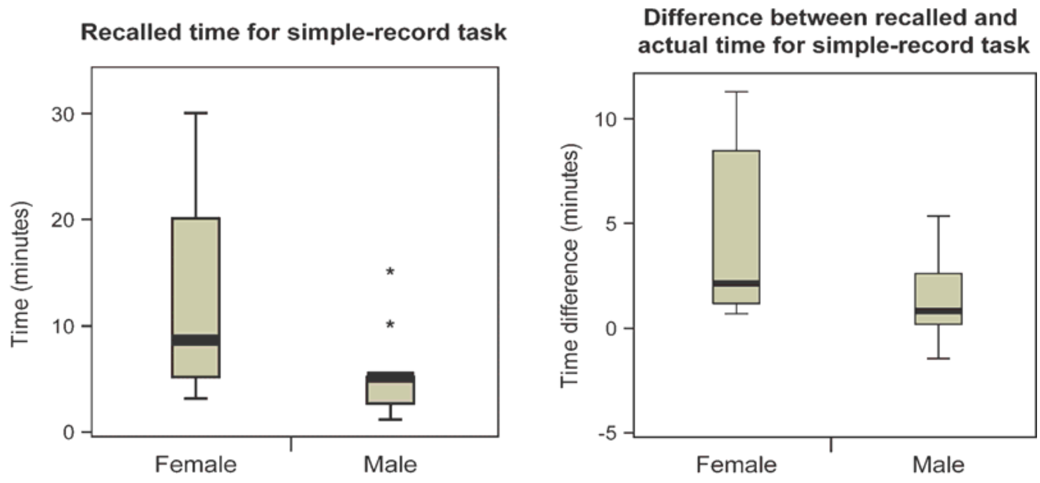

Figure 6. Participant's estimates of the time that they had taken in the simple-record task: a) female participants recall the task as having taken longer than males do; b) females recall is more inaccurate, relative to their actual performance. One extreme value (female estimating over two hours) has not been shown, in order to improve visual clarity on the vertical axis, but all data was included in statistical analysis.

As shown in Figure 6a), females recall the simple task as having taken longer than males do (Mann Whitney U, $p<.05$). To some extent this is simply realistic, because as noted, females did take longer on all tasks. However, we can assess the degree of realism by comparing the *recalled* time estimate to the *actual* time taken, as in Figure 6b). On average, females estimated that the simple-record task had taken 4 or 5 minutes longer than it actually had taken them (in fact, an average of 10 minutes longer, if the most extreme outlying overestimate is included). Males also overestimated the time taken in the simple-record task, but by a far smaller amount (Mann Whitney U, $p<.05$). There was also a tendency for females to recall the program-ahead task as having taken longer than it really did, although this difference between male and female participants was not significant (Mann Whitney U, $p=.10$).

With regard to estimated likelihood of success after the task had been completed, we observed that females tended to revise their estimated likelihood of success downward after completing the program-ahead task, while males tended to revise their estimate upward. This difference in estimated likelihood of success was not statistically significant, but it would be consistent with the tendency in estimating difficulty, in that females revised estimates of both difficulty and risk of failure in a direction that is likely to bias them against future programming. Qualitative analysis of our questionnaire data provided independent support of this observation, in that females who carried out the simple-record task first described themselves in more confident terms than those who carried out the program-ahead task first. The latter more often described uncertainty: "I wasn't sure", "I was initially scared", "I'm not used to doing this", "what else could it be" or "it knocks your confidence". These observations are consistent with recent research into gender effects in end-user web design (Rosson et al. 2007), which has found that males rated themselves more highly than females with regard to "perceived success" after the programming task was completed.

Post-hoc test: Compensation for Gender Biases in Experience

Overall, these results tend to confirm the findings of Beckwith, that there are significant gender effects in the factors that influence strategy choice and confidence of task completion. One intervention strategy proposed by Beckwith is that female users can positively revise their estimation bias toward programming strategies if they have some experience of "tinkering" with the programming system. We made a post-hoc test for any effect of tinkering in our results (post-hoc because this experiment was designed before Beckwith's results were formulated). "Tinkering", as defined by Beckwith, refers to gaining experience by use of simpler software functions, that build user understanding and confidence toward attempting larger programming tasks. In the context of our experiment, we consider the simple record task to offer this kind of confidence-building experience, although it should be noted that Beckwith's results referred to repeated use of simple functions, whereas our participants only completed the simple-record task once.

A positive effect similar to "tinkering" in our experiment might have been indicated by an increase in confidence after successful completion of the simple-record task. Repeated use of this simple function over time could then build further confidence. Unfortunately our measures were not significantly sensitive to find a significant effect after a single task (further research might perhaps use measures such as self-efficacy questionnaires, as well as the opportunity to repeat simple tasks). However there is some ground for doubt regarding any "tinkering" benefit, judging by the fact that female participants over-estimated the time they took to complete the simple record task. This suggests that, in order to create a beneficial tinkering effect for female users, the operations concerned should be a lot more straightforward than our simple-record task. We would also expect it to be counter-productive if supposedly simple operations were accompanied by breakdowns like the problems that our participants had in gaining access to the onscreen menu, as these can easily reduce user confidence rather than increase it.

Implications for Design

Our objective in this research has been to better inform the design of programmable home appliances. The Attention Investment model of abstraction use describes programmable functions by contrast to direct manipulation functions. In direct manipulation, the effects of user actions take place immediately, with immediate feedback. Our research in homes and retail stores confirms that there are an increasing number of products offering features where this is not the case. Abstract product functions allow users to change the behaviour of the product in future. In the relatively simple cases that we have described, the future changes are restricted to a small class of behaviour. However, this need not be true in future, and indeed is unlikely to remain true. New classes of product such as personal video recorders and personal MP3 players have brought new types of configurable abstraction (personal viewing preferences and playlists respectively). Rapid increase in domestic wireless networking and internet appliances will support further opportunities for programming, especially where there is a requirement to define the interaction between different products in one's own house.

Value of the Attention Investment Model

Although the range of abstract and programmable functions is increasing, designers might ask whether there is any underlying commonality between this wide range of features. Perhaps each new feature should be designed to meet the specific user need it anticipates, possibly adopting proven design conventions from computer user interfaces when they are directly relevant (e.g. using icons and menus)? As a theoretical base for design, the Attention Investment model of abstraction use offers a stronger alternative. We have demonstrated that a model originally motivated by psychology of programming, which has previously been shown to apply to word processor, file management, and spreadsheet use, applies also to programmable appliances in the home. The Attention Investment model is both descriptive and predictive. The descriptive contribution is to remind designers that many users choose not to use abstract functions, not because those users are less intelligent than early adopters and other “geeks” (possibly including the designer), but because users make rational choice models based on expected costs, benefits and risks. If users fail to employ a feature in the most appropriate way, the designer must take responsibility for failing to communicate the benefits (or costs or risks) appropriately.

The predictive contribution of the Attention Investment model is in design contexts where it is necessary to anticipate user response to new features, for example in the more conventional end-user programming domain of spreadsheet use. As observed in the work of Burnett and collaborators in the EUSES consortium for end-user software engineering, spreadsheet users often fail to invest attention in improving the quality of their spreadsheets. It is possible to implement sophisticated features for testing and specification of spreadsheets, but commercial spreadsheets already have many features, and users are unwilling to invest the attention in learning to use more of them. Application of the Attention Investment model suggested specific design interventions that would encourage users to employ features that would improve spreadsheet quality. As described above, the model predicted that users would not investigate new features unless their existing biases toward habitual strategies were disrupted in some way, and that new features would not be employed unless users could understand and verify the benefits in terms of attention costs and benefits. This led to the design strategy of “surprise, explain, reward”, in which users were prompted to reassess their attention investment judgments (Wilson et al. 2003).

Design to Account for Gender Biases in Estimation

In demonstrating the ways that the Attention Investment model might also influence strategy choice in use of abstract functions for configuring and controlling domestic appliances, we see an opportunity for similar design strategies in this domain. One opportunity to which we have paid particular attention is the existence of gender biases in strategy choice. We showed that, for a particular class of appliance, interaction between self-efficacy and the Attention Investment model predicts that females are less likely to choose a programming strategy – because they expect programming to involve more effort, with less likelihood of success, than males do. Furthermore, even after successful completion of typical tasks, females underestimated their likelihood of success and overestimate the effort that had been involved, making them less likely to employ apparently riskier, but more profitable, strategies in future. Risk, in a broader analysis of the domestic environment, might also take into account “social risks”, such as altered perception of a user’s femininity, or adverse consequences for domestic harmony.

We believe that some straightforward design manoeuvres could make abstract features more accessible. One is to incorporate opportunities for “tinkering” like operation, which should build confidence in operation of the appliance, ideally including experience of those interface features that will recur in the context of the abstract task. Another is to avoid initial experiences that might bias users against further abstract strategies. This appears to have been particularly problematic for female users, who are just as competent, but interpret their experiences differently, leading to attention investment estimation biases against future programming strategies.

Interaction Between Appliances

Although this experiment focused on abstract control of a single appliance, our long-term goals were motivated by research into networked home technologies, where increasingly complex abstractions will be required to specify and control the potential interactions between networked appliances. The Attention Investment model of abstraction use describes the likely transfer of strategies from programming of stand-alone appliances to these more complex programming tasks. Ironically, although network interaction was not the focus of this particular experiment, it still arose as a factor even in our simple scenario. Our experimental setup involved two appliances: the DVD recorder, and the TV on which recorded programmes would be viewed. As has been typical in this generation of video technology, much of the user interface for the DVD recorder is provided as on-screen menus on the TV. This requires some degree of inter-operability between the two, albeit far less than will be required in future generations of networked appliances.

While we knew based on our previous consumer electronics experience that inter-operability between devices may pose a challenge for participants in this experiment we had attempted to mitigate this problem. We bought these two devices on the same day, from a retail store well regarded for its expert customer advice, and the two were made by the same manufacturer. They were both in the mid-price range for their categories at the time (DVD recorders have since become less expensive), and could be regarded as the standard consumer offering from that manufacturer. Some inter-operability between the two was clearly assumed by the manufacturer, because the remote controls offered some support for cross-control of functions. Despite all of these factors, the single greatest usability challenge faced by our participants was managing the interaction between the two products, and especially the need to access the control functions of the DVD recorder via the TV screen. In fact, this repeated an observation from our diary study of the domestic programming context, where we had monitored any task failures during the two week period. The one failure to replay a recording resulted from the same factor seen here – a member of the household was unable to determine how to display the VCR menu on the TV screen. Although not directly related to our theoretical motivations in Attention Investment and gender (other than the effect that it has on building confidence), this observation should be noted as a significant design priority.

Market Relevance

We recommend that attention investment analysis can alert designers to the way in which programmable features might be adopted and integrated into documentation, menu structures and control panels. In particular, past failures of “intelligent” strategies for user assistance (e.g. the Microsoft office assistant “Clippy”, or in the home media domain, the “TiVo Suggestion” functions) might be analysed according to their attention investment characteristics. At the time of writing, some cross platform initiatives (such as Universal Plug and Play) are not engaging with issues of network programmability and customisation. Manufacturers such as Bang and Olufsen have businesses that were historically built on interoperability and networking, but are now struggling to retain the advantage of specialist high-end services. New technology players such as Nokia and Samsung are trying to exploit the potential of their mobile computing platforms as universal home controllers. Meanwhile, new start-ups are creating disruptive integration technologies that are being deployed in the home for single-solution purposes (e.g. home security), but with strategic capability to expand their ownership of the home control market into major cross-media businesses. At this point, the technical decisions and alliances being made will have critical consequences for users of next generation technologies. It may be the case that configuration of domestic technology will be centralised or professionalised (e.g. Spinellis 2003). However we believe that this design approach carries with it social and gender implications. We hope that the “end-user programming” ethos might be established among the designers of domestic technology, not simply by association with the hobbyist market (as in previous generation of X10 home automation), but through the use of an Attention Investment model to anticipate the biases associated with rational strategic choices.

Acknowledgments

Funding for the retail survey of consumer electronics and furnishing of the usability lab was provided by the Cambridge-MIT Institute. Funding for the ethnographic study of home programming, diary study of VCR usage, and experimental study of DVD programming was provided by EPSRC grant GR/R87482/01, *Cognitive Ergonomics for Ubiquitous Computing*. Alan Blackwell’s research is also supported by the National Science Foundation grant ITR-0325273, *End-Users Shaping Effective Software (EUSES)*.

References

Antonides, G., and van Raaij W.F. (1998). *Consumer behaviour: a European perspective*. Chichester, Wiley.

Bandura, A. (1986). *Social foundations of thought and action*. Prentice Hall, Englewood Cliffs, NJ.

Beckwith, L., Kissinger, C., Burnett, B., Wiedenbeck, S., Lawrance, J., Blackwell, A. and Cook, C. (2006). Tinkering and gender in end-user programmers' debugging. In *Proceedings of CHI 2006*, pp. 231-240.

Berk, S.F. (1985). *The Gender Factory: The Appointment of Work in American households*. New York, Plenum.

Blackwell, A.F. (2002). First steps in programming: A rationale for Attention Investment models. In *Proceedings of the IEEE Symposia on Human-Centric Computing Languages and Environments*, pp. 2-10.

Blackwell, A.F. (2004). End user developers at home. *Communications of the ACM* 47(9), 65-66.

Blackwell, A.F. (2006). Psychological issues in end-user programming. In H. Lieberman, F. Paterno and V. Wulf (Eds.), *End User Development*. Dordrecht: Springer, pp. 9-30.

Blackwell, A.F. and Hague, R. (2001). AutoHAN: An architecture for programming the home. In *Proceedings of the IEEE Symposia on Human-Centric Computing Languages and Environments*, pp. 150-157.

Blackwell, A.F., Hewson, R.L. and Green, T.R.G. (2003) Product design to support user abstractions. In E. Hollnagel (Ed.) *Handbook of Cognitive Task Design*. Lawrence Erlbaum Associates, pp. 525-545.

Blackwell, A.F., Rode, J.A. and Toye, E.F. (2005). The social context of domestic end-user programming. In *Proceedings of Less is More*, Cambridge UK.

Brumitt, B., and Cadiz, J.J. (2001). *Let there be light" examining interfaces for homes of the future*. INTERACT'01. IFIP TC.13 International Conference on Human Computer Interaction., Amsterdam, Netherlands, IOS Press. 375-82.

Bureau of Labor Statistics (2004). *Employment status of the civilian noninstitutional population 16 years and over by sex, 1970-2004 annual averages* <http://www.bls.gov/cps/wlf-table2-2005.pdf> (Last accessed November 26, 2007).

Carroll, J.M. and Rosson, M.B. (1987). Paradox of the active user. In J.M. Carroll (Ed.), *Interfacing Thought: Cognitive Aspects of Human-Computer Interaction*, Bradford Books/MIT Press, pp. 80-111.

Consolvo, S., Roessler, P. and Shelton, B.E. (2004). The CareNet Display: lessons learned from an in home evaluation of an ambient display. In Proc. *Ubicomp*. 1-17.

Deem, R. (1986). *All Work and No Play? The sociology of women and leisure*. Milton Keynes, OU Press.

Gigerenzer, G., and Selten, R. (Eds.) (2001). *Bounded rationality: the adaptive toolbox.* Cambridge MA: MIT Press.

Grinter, R.E., Ducheneaut, N., Edwards, W.K. and Newman, M. (2005). The work to make a home network work. In *Proc. 9th European Conference on Computer Supported Cooperative Work (ESCW '05),* pp. 469-488.

Harper, R. (Ed.) (2003). *Inside the smart home*. London: Springer Verlag.

Isbell, C.L. Jr., Omojokun, O. and Pierce, J.S. (2004) From devices to tasks: automatic task prediction for personalized appliance control. *2nd International Conference on Appliance Design*, pages 23-29.

Livingstone, S.M. (1992). The meaning of domestic technologies: a personal construct analysis of familial gender relations. In R. Silverstone and E. Hirsch (Eds).

Consuming technologies: Media and information in domestic spaces. London: Routledge, pp. 113-130.

MacLean, A., Carter, K., Lövstrand, L. and Moran, T. (1990). User-tailorable systems: pressing the issue with buttons. In *Proceedings of the Conference on Human Factors in Computing Systems – CHI'90*, pp. 175-182.

Maushart, S. (2001). *Wifework; what marriage really means for women*. New York, NY, Bloomsbury.

Mynatt, E.D., Essa, I. and Rogers, W. (2000). Increasing the opportunities for aging in place. *Proceedings on the 2000 conference on Universal Usability*: 65-71.

Pane, J.F. and Myers, B.A. (2006). Natural programming languages and environments. In H. Lieberman, F. Paterno and V. Wulf (Eds.), *End User Development*. Dordrecht: Springer, pp. 31-50.

Potter, R. (1993). Just-in-time programming. In A. Cypher, (Ed.), *Watch what I do: programming by demonstration*. Cambridge, MA: MIT Press, pp. 513-526.

Rode, J.A., Toye, E.F. and Blackwell, A.F. (2004). The Fuzzy Felt Ethnography - understanding the programming patterns of domestic appliances. *Personal and Ubiquitous Computing* 8, 161-176.

Rode, J.A., Toye, E.F. and Blackwell, A.F. (2005). The domestic economy: A broader unit of analysis for end user programming. In *proceedings CHI'05 (extended abstracts),* pp. 1757-1760.

Rosson, M.B. Sinha, H., Bhattacharya, M. and Zhao, D. (2007). Design planning in end-user web development. In *Proc. IEEE Symposium on Visual Languages and Human-Centric Computing*, pp. 189-196.

Russell, S. and Wefald, E.H. (1991). *Do the Right Thing: Studies in limited rationality*. Cambridge, MA: MIT Press.

Simon, H. (1956). Rational choice and the structure of the environment. *Psychological Review* 63, 129-138.

Sohn, T.Y. and Dey, A.K. (2003). iCAP: An Informal Tool for Interactive Prototyping of Context-Aware Applications. In *Proceedings CHI'03 (extended abstracts),* pp. 974-975.

Spinellis, D. (2003). The Information Furnace: consolidate home control. *Personal and Ubiquitous Computing* 7: 53-69.

Staton, S. (2005). *An agent architecture for simulation of end-users in programming-like tasks*. University of Cambridge Computer Laboratory, Technical Report UCAM-CL-TR-647.

Wilson, A., Burnett, M., Beckwith, L., Granatir, O., Casburn, L., Cook, C., Durham, M. and Rothermel, G. (2003). Harnessing curiosity to increase correctness in end-user programming, In *Proc. ACM Conf. Human Factors in Computing Systems (CHI'03),* pp. 305-312

Appendix A: Pre-task interview script

'Have you used a DVD recorder (not just a DVD player) before?'
□ Yes □ No

IF YES: 'Have you used a Panasonic DVD recorder before?'
□ Yes □ No

IF YES: 'Have you used this particular model of Panasonic DVD recorder before?'
□ Yes □ No

'Approximately how many minutes do you think it will take you to do this task?'

□ 5 mins or less □ 10 mins or less □ 15 mins or less
□ 30 mins or less □ 45 mins or less □ 1 hour or less
□ More than 1 hour

'How easy or difficult do you think the task will be?'

1	2	3	4	5	6	7	8	9
Very difficult		Difficult		Neutral		Easy		Very easy

'How confident are you that you will be able to complete the task successfully?'

1	2	3	4	5	6	7	8	9
Very Un-confident		Un-Confident		Neutral		Confident		Very Confident

'If you get stuck, please make a reasonable effort to solve the problem – but if you can't make any further progress, you can give up. If you need to take a break for any reason, that's ok too, just let us know and we'll stop the clock until you're ready to start again.

Do you have any questions before you start?'

Questions about the task

'How long do you think it took you to complete the task?'

□ 5 mins or less □ 10 mins or less □ 15 mins or less
□ 30 mins or less □ 45 mins or less □ 1 hour or less
□ More than 1 hour

'How easy or difficult did you find the task?'

1	2	3	4	5	6	7	8	9
Very difficult		Difficult		Neutral		Easy		Very easy

'How confident are you that you were successful in setting up the DVD-R to record the programme I asked you to record?'

1	2	3	4	5	6	7	8	9
Very Un-confident		Un-Confident		Neutral		Confident		Very Confident

If yes: ‘Why do you think you were successful?’
If no: ‘Why do you think you were unsuccessful?’
If not sure: ‘Why are you unsure whether you were successful?’

‘What factors made the task easier for you? What factors made the task more difficult for you?’

Appendix B: Post-task interview script

Does your household own any of the following appliances? (Tick all that apply)

□ DVD recorder
□ DVD player (no recording facility)
□ Video recorder
□ Integrated TV/video recorder
□ Sky Plus
□ TiVo
□ Other TV recording appliance (If ‘other’ please tell us what it is!)

What experience do you *personally* have of using a:	Never	Playback, but not record	Record by pressing ‘Record’	Record – set up ahead, using written help	Record – set up ahead, w/o written help
DVD recorder	(before today)				
DVD player (no record function)			N/A	N/A	N/A
VCR					
TV/VCR combo					
Sky Plus					
TiVo					
Other TV recording appliance					

How often do you *personally* use a:	Never	Very rarely	More than once a year	More than once a month	More than once a week
DVD recorder	(before today)				
DVD player (no record function)					
VCR					
TV/VCR combo					
Sky Plus					

TiVo					
Other TV recording appliance					

'Can I ask you how confident you usually are about using technical equipment?'

Which of the following best describes your technical ability?

1. very bad at technical tasks.
2. fairly bad at technical tasks.
3. of average ability at technical tasks.
4. fairly good at technical tasks.
5. very good at technical tasks.

'How confident are you about pre-recording TV programmes?'

Which of the following best describes your ability to pre-record TV programs?

1. very bad at pre-recording TV programmes.
2. fairly bad at pre-recording TV programmes.
3. of average ability at pre-recording TV programmes.
4. fairly good at pre-recording TV programmes.
5. very good at pre-recording TV programmes.

'Besides TV recording equipment, could you tell me what other types of domestic technology you own or use regularly?

'What do you think about the design of TV recording equipment, from a user's viewpoint, compared to other domestic technology you are familiar with?'

'Which of the following best describes how you would compare the design of TV recording equipment to other domestic technology?'

1. much worse designed than other domestic technology.
2. somewhat worse designed than other domestic technology.
3. no better or worse designed than most other domestic technology.
4. somewhat better designed than other domestic technology.
5. much better designed than other domestic technology.

'From a user's viewpoint, what do you think about the design of the DVD recorder that you learned to use today?'

Which of the following best describes how you would compare the design of this DVD-recorder to other TV recording equipment?'

1. much worse designed than most TV recording equipment.
2. somewhat worse designed than most TV recording equipment.
3. neither better nor worse designed than most TV recording equipment.
4. somewhat better designed than most TV recording equipment.
5. much better designed than most TV recording equipment.

'Do you have any disabilities or health problems which might affect your ability to use TV recording equipment? If so, could you tell me about that?'

'Do you do any computer programming?'