Interleaving Tasks to Improve Performance: Users Maximise the Marginal Rate of Return

Geoffrey B. Duggan[1], Hilary Johnson[2] and Petter Sørli[2]

[1]University of Exeter

Exeter, United Kingdom

[2]University of Bath,

Bath, United Kingdom

Correspondence concerning this article should be sent to: Geoffrey Duggan,

Psychology, College of Life and Environmental Sciences, University of Exeter, Exeter, EX4 4QG, UK. Email: gd268@exeter.ac.uk

Abstract

Technological developments have increased the opportunity for interleaving between tasks, leading to more interruptions and more choices for users. Three experiments tested the interleaving strategies of users completing simple office-based tasks while adjusting access control privileges to documents. Previous work predicted users would switch tasks to enable them to work on the task that produced the greatest current benefit – they would maximise the marginal rate of return. Results found that by interleaving between tasks users were able to focus on shorter tasks and that the interleaving decisions were consistent with a strategy of maximising the marginal rate of return. However, interruptions from access control tasks disrupted the processing involved in this task management and led to errors in task selection (Experiment 2) and task performance (Experiment 3). Task interleaving can therefore have costs in security contexts where errors can be catastrophic. Understanding which strategies maximise the marginal rate of return could predict users' task management behaviour.

Keywords: Interleaving, Multitasking, Interruption, Task management, Discretionary access control

1. Introduction

Technology enables people to contact one another, to access vast depositaries of information and to change tasks all with the press of a button. These capabilities also mean that workers have to cope with frequent interruptions and switches between tasks (Czerwinski et al., 2004; Gonzalez and Mark, 2004; McFarlane and Latorella, 2002; Wild et al., 2003). This article studies the way that users manage interruptions and task switches and select between competing task demands. Without an account of how or why users choose to change tasks it is difficult to exploit this ubiquitous behaviour and mitigate the costs of stopping work on unfinished tasks.

Research into task interruption within human-computer interaction (Adamczyk and Bailey, 2004; Bailey and Konstan, 2006) and psychology (Cades et al., 2011; Gillie and Broadbent, 1989) typically investigates the effect of an external interrupting stimulus upon performance on a single task. Under these conditions an interruption has been shown to degrade primary task performance (Altmann and Trafton, 2002; Hodgetts and Jones, 2006), although for simple tasks interruptions can sometimes be helpful (Speier et al., 1999; Speier et al., 2003). Less understood is the metacognitive process of task management. That is, where there are multiple tasks to be completed, users much select which task to work on when, and account for the impact of interruptions on task management as well as the impact of interruptions on the tasks themselves. Task management has a large impact on worker productivity (Jackson et al., 2001; Mangelsdorf, 2011) and our experiments are designed to understand the process of task management.

Task management processes include both internally initiated "self-interruptions" and interruptions from external stimuli (Benbunan-Fich et al., 2011). Gonzalez and Mark (2004) and Czerwinski et al. (2004) found that almost 50% of task switches in office environments were initiated by the user, furthermore, self-interrupted tasks are less likely to be resumed than tasks interrupted by external events (Mark et al., 2005). Thus, unlike most studies of task interruption, we shall focus on users' self-interruptions. We also investigate the effect of external interruptions but upon task selection processes rather than primary task performance.

Task management processes have been studied within research into multitasking and task switching (e.g. Rogers and Monsell, 1995, Rubinstein et al., 2001). This work looks at the simultaneous completion of relatively short tasks and the bottlenecks in cognitive processing that result. Controlled laboratory studies have developed sophisticated theories of the allocation of resources to tasks that are completed in parallel (Salvucci and Taatgen, 2008). Furthermore, these bottlenecks can have important consequences for performance on tasks such as dialling telephone numbers while driving (Brumby et al., 2009). However, this work describes very different task management processes to those reported by researchers observing office workers *in situ*. Our interest is in understanding the switches between tasks that on average are made by office workers every 4 to 11 minutes (Gonzalez and Mark, 2004; Mark et al., 2005). Therefore, we shall focus on the selection between tasks or subtasks that are worked on serially until some other task is chosen. At this level, effective task management relies on the prioritisation of the most important task rather than the characteristics of temporary resource bottlenecks. Following Payne et al. (2007) we refer to these task management processes as *discretionary task interleaving*.

1.1. Discretionary access control and discretionary task interleaving

The way that users manage discretionary task interleaving has implications for security. As more information is stored and exchanged digitally, there is an increasing need to control access to this information. Making digital resources more convenient and accessible can mean greater vulnerability to security threats (Beautement et al., 2008). Effective solutions to this problem must account for the people who use the technology and implement security policy (Herley, 2009). For example, traditional approaches to access control such as Role-Based Access Control (Sandhu et al., 1996) are effective where administrators can allocate privileges in a systematic way. However, users increasingly share information through informal structures such as social networking sites (Bradwell and Reeves, 2008) meaning the individual user has greater discretion over who has access to what information (Carminati et al., 2007).

This shift towards *discretionary access control* means that task management processes directly affect the security of information stored digitally. Switching out of an incomplete task or forgetting to start/return to a task could prevent authorised people from accessing information or enable unauthorised people to access sensitive information. Furthermore, the security literature indicates that user errors are commonplace in real-world use of access control systems (Smetters and Good, 2009) and more generally user errors are a primary cause of breaches of security (Whitman, 2003).

Another relevant feature of access control tasks is that they are typically enabling tasks. Setting access privileges, requesting access and waiting for access to be granted are all tasks that facilitate another primary task (e.g. reading the document to which access is requested). This means that access control is often interleaved with other tasks and can act as an interruption to the users' primary task (Cao and Iverson, 2006). In summary, access control tasks are increasingly widespread and users have greater discretion over their completion. Access control is frequently interleaved with other tasks but errors in task management can have important consequences. As such, access control provides an ideal domain in which to study discretionary task interleaving.

1.2. Discretionary task interleaving

The fundamental goal of task management is to allocate time in such a way as to achieve the most important goals. Interleaving between tasks is a method to account for changes in the current task, the status of alternative tasks and the goals of the user. A fuller account of the reasons *why* users self-interrupt or interleave between tasks was provided by Jin and Dabbish (2009). They identified seven different categories of self-interruption, these included types that were initiated by factors internal to the user (e.g. breaks due to boredom/fatigue or recollections of other incomplete tasks) and those initiated by external situational factors (e.g. an adjustment to the environment to aid productivity or an enforced wait caused by a temporary roadblock in the current task). Interleaving is therefore a commonplace activity (Gonzalez and Mark, 2004) with many different purposes (Jin and Dabbish, 2009) but little is known about its effectiveness. The current experiments investigate whether users are able to

manage tasks with competing demands by interleaving between them and how users decide when and if to interleave between tasks.

Payne et al. (2007) addressed these questions by giving participants two sets of letters and asking them to generate as many words as possible within a fixed period of time. Participants were allowed to interleave between letter sets, but a word could not use letters from both sets. Results were that participants interleaved between the letter sets throughout the time period and overall spent more time on the easier letter set that contained more words. Interleaving enabled participants to maximise the total number of words generated. However, participants could have achieved this with far fewer switches between letter sets – on average they switched seven times in ten minutes. Their behaviour was consistent with a strategy of maximising the marginal rate of word generation. That is, they worked on the task that gave the greatest rewards at that point in time. Because there were diminishing marginal returns in both letter sets (it got harder to generate new words over time) this meant the task with the highest marginal rate of return changed over time leading to switches between tasks.

Pirolli (2007) has drawn parallels between animal foraging and human information foraging. Similarly, we note that the marginal rate of return provides a good account of switches between food patches by foraging animals (Charnov, 1976) and switches between sections of text by skim readers (Duggan and Payne, 2009, 2011). We suggest that maximising the marginal rate of return could extend beyond the experiments in Payne et al. (2007) and explain the frequent self-interruptions by office workers (Czerwinski et al., 2004; Gonzalez and Mark, 2004). Task rewards often vary over time as does the availability of alternative tasks meaning that the task with the highest marginal rate of return will often change over time.

In a typical office situation workers must monitor the marginal rate of return on several different types of task with many distractions, whereas Payne et al. (2007) only considered interleaving between very simple word generation tasks. When individuals must choose between different types of task or manage distractions a key part of the problem is converting performance across tasks to some common task currency. It is difficult to compare rates of return when they are in different currencies. Our experiments incorporate this need to convert

task currencies by including tasks more representative of the workplace within a complex environment containing distractions and different types of tasks. Thus, we move a step closer than Payne et al. (2007) to everyday discretionary task interleaving. In Experiments 1 and 2, we specify that the goal for participants is to complete as many tasks as possible within the time provided and assess whether they interleave between tasks to maximise the marginal rate of return. In Experiment 3, we use the marginal rate of return to understand errors in task performance.

1.3. Interruptions to task management processes

The interference caused by interruptions to single task performance (Altmann and Trafton, 2002; Hodgetts and Jones, 2006) suggests the metacognitive processes used for task management will also be degraded by interruptions. However, Mark et al (2008) interrupted participants completing multiple tasks and found that the interrupted tasks were completed in less time and with no difference in quality relative to uninterrupted tasks. Mark et al. (2008) suggested that participants compensated for the interruptions by working harder on the primary task. However, it is difficult to discern the specific effect of the interruptions on task management as performance was assessed using aggregate measures that included both primary task performance and task management. To address this issue we shall directly measure the task management decisions.

Given that task management is an ongoing activity it is not straightforward to determine what should be classed as an interruption to the processing involved. Nonetheless, in Experiments 1 and 2, task management was assessed using decisions about the primary task and the access control tasks were additional tasks to be completed. Therefore, we shall consider access control tasks as interruptions to task management and assess the effect of these interruptions by measuring task selection time within the primary task. Experiment 3 provides a different demonstration of the effects of interruptions on task management by comparing task performance when participants do and do not receive alerts for email interruptions.

2. Experiment 1

To test discretionary task interleaving in a more realistic environment, typical office based tasks were constructed that reproduced key elements of Payne et al.'s (2007) experiments. Specifically, a set of document editing tasks of varying lengths were divided into an easy folder (containing a high proportion short tasks) and a hard folder (containing a high proportion of long tasks). Participants were instructed to complete as many tasks as possible within the time provided, they were allowed to interleave between folders as much or as little as they chose and the rate at which they completed tasks acted as the rate of return. Within each folder, tasks were ordered from shortest to longest meaning that there was a diminishing marginal rate of return within each folder. Following Payne et al. (2007), we predict that participants will interleave between the folders throughout the experiment to maximise the marginal rate of return and they will complete more tasks from the easy folder than the hard folder.

In addition to the document editing tasks one condition also had to complete access control tasks. These used the Information Rights Management (IRM) system from Microsoft Office 2010. The access control tasks required participants to set access privileges for the document after completing a document editing task and to request permission to open documents for which they did not have access. Participants had to interleave between these tasks and the document editing tasks. We predict that time to select the next document-editing task will be longer when participants are also required to complete access control tasks.

Although access control tasks are increasingly familiar, their use within Microsoft Office was relatively new when the experiments were conducted. Therefore, the third objective was to investigate user performance on the access control tasks when interleaving them with other primary tasks.

The hypotheses are:

H_1: Participants will complete more tasks from the easy folder than the hard folder.

H_2: Task completion time will be longer prior to a switch between folders than after a switch between folders.

H_3: Task selection time will be longer in the IRM condition than the Control condition.

2.1. Method

2.1.1. Participants

Participants were 13 students and 11 staff from a research based university. They received £10 for taking part in the experiment. Thirteen participants were female and the age range was 19-56 years ($M = 31.75$, $SD = 9.47$). All participants reported they were familiar with Microsoft Office and had used it for more than 10 hours.

2.1.2. Materials

A set of 100 documents, one page in length were constructed from actual documents found at www.mrsc.org/sampledocs.aspx. For each document a task was devised that required participants to add text at a specific point in the document or as a replacement for text within the document (e.g. "Replace January 1, 2004 with November 1, 2007"). The amount of text that had to be added varied across tasks from 1 word to 124 words. The length of time to complete each task was measured using the number of characters of text that had to be entered. Tasks were divided into lists of 25 tasks, 2 lists for the easy folders and 2 lists for the hard folders. Each list was ordered from shortest task at the top to longest at the bottom and each task in an easy list was matched in length to the corresponding task in the other easy list. The hard lists were matched in the same way. Tasks were matched according to character length then one task was randomly allocated to a list and the matched task was allocated to the corresponding list. Figure 1 shows the length of each task for each of the 4 lists.

Materials were presented using a program purpose-written for the experiment. Each list of documents was contained in a separate folder, labelled either Folder 1 or Folder 2, which when opened filled the screen. Clicking on a document caused it to be opened in a simulation of Microsoft Word. The folder interface was the standard navigation interface within Windows and the interface for sending and receiving emails was taken from the University's webmail package. Tabs corresponding to each program were provided at the bottom of each screen and clicking on them was the only way to navigate between the word simulation,

email and different folders. Each window was locked so that it could not be altered in size or minimised, only one document could be opened at a time. A screenshot from the word simulation interface is given in Figure 2.

Documents were presented in a textbox superimposed on the word simulation interface, the textbox allowed participants to type, delete, cut, copy and paste text. All other features contained with the original Microsoft Word were disabled. The task was presented in a window to the right of the document (see Figure 2), text could not be cut or copied from the task window to the document and had to be entered using the keyboard. Within each folder the top two documents were labelled in red indicating they were priority documents. These documents also contained the shortest tasks in each folder. After a document had been opened within the folder it was labelled in blue. Within the email simulation participants could write messages, check for new mail and read mail in the same way as normal within the webmail package but all other features were disabled. Email did not automatically update the inbox when it arrived and, as in the actual package, participants had to refresh the screen by clicking on the "Inbox" icon. When sending messages there was no access to an address book meaning that participants had to enter the recipient's address using the keyboard.

The procedure for adding access rights in Word 2010 was also simulated within the program. After clicking on "File" participants had to complete a five step procedure: Step 1, Click on Protect; Step 2, Select "Restricted Access" from the menu; Step 3, Click on OK within the dialog box; Step 4, Check the box next to "Restrict permission to this document"; Step 5, Enter the email address of the person to be given access and click on OK. The email address of the person to be given access was provided in a box in the corner of the screen, this address was different for every document. Apart from the option to return to the document interface all other features were disabled within the word simulation thus the program did not change state after any errors.

In the IRM condition access was denied to some documents. After clicking on these documents a dialog box appeared that explained that access was denied and gave the name of the person who should be emailed to gain access. Throughout the experiment a clock in the bottom right hand corner of the screen counted down the number of seconds remaining in that

condition. The standard clock within Microsoft Office was not used as it did not display seconds. All interactions with the program were recorded and timestamped.

2.1.3. Design

Independent variables were Experimental Condition (IRM vs. Control) and Task Difficulty (Easy Folder vs. Hard Folder), both were manipulated within participants. The order in which the conditions were completed and the position of the Easy and Hard folders in the interface as either "Folder 1" and "Folder 2" were counterbalanced across participants. Task completion time was measured from the selection of a document from a folder to the decision to navigate away from the document interface (either by opening a folder or email package or by starting the access control procedure).

Task selection time was measured from the opening of a folder following a task to the next time a document was opened. For all measures of task selection time, when participants were interrupted in the IRM condition (i.e. denied access to a document, used the email package or interleaved between these activities and a folder), the length of the interruption was subtracted from the task selection time. An additional measure of uninterrupted task selection time was also computed which completely excluded any between task intervals where the email package was opened or access to a document was denied.

2.1.4. Procedure

Before each condition participants were given written instructions explaining the program and the tasks that had to be completed. These instructions stated that the overall goal was to complete as many tasks as possible within the time period. The experiment was conducted in a laboratory, the experimenter and the participant and no other persons were present throughout the experiment. Three practice tasks were completed first by the experimenter and then by the participant. The practice tasks differed according to the experimental condition. For the Control condition, all 3 tasks were in the same folder and participants simply had to complete the tasks in any order they chose. They achieved this by clicking on a document, making the changes necessary to the document and then clicking on the folder to select

another task. The program automatically saved all changes to documents and it was not necessary to save any changes at any point.

In the practice tasks for the IRM condition participants had to complete the 3 tasks in the same way as in the Control condition with two additional components. Firstly, after completing each task they had to restrict access to the document to a particular person by following the 5-step procedure described in the Materials section. Secondly, participants were denied access to one of the documents and had to gain access to it. The dialog box denying access to a document also provided the email address of the document owner. Participants were then required to write a short email to this address requesting access to the document. A reply was received in 5 seconds granting this request and the document could then be opened as usual.

After completing the first set of 3 practice tasks, participants then completed the corresponding experimental condition followed by a 2 minute break, the practice tasks and then the other experimental condition. For the experimental conditions participants were given access to a folder containing an easy list of documents and another folder containing a hard list of documents. They were not instructed that the folders differed in task composition, but were told that within each folder the documents were organised such that documents with shorter tasks were at the top of the folder and documents with longer tasks were at the bottom of the folder. They were free to select tasks in any order they chose but were required to complete as many tasks as possible within the time limit provided.

They were also instructed that at the top of each folder were 2 tasks labelled in red, these were “Priority tasks” and if possible should be completed before other tasks. The experimental conditions took the same format as the practice trials meaning that in the Control condition, participants simply had to select tasks and complete them, whereas in the IRM condition, participants were also instructed to restrict access to every document using the 5-step procedure and were denied access to some documents. They were denied access to the second document in each folder and the eighth document in the easy folder. The reply to an email requesting access was sent after 30s for the second document in the easy folder, after 240s for the second document in the hard folder and after 60s for the eighth document in

the easy folder. If an email was sent to the wrong address participants did not receive a reply. Participants were instructed that they could work on other tasks while waiting for a reply to an email.

The total time for task completion was 2200s in the IRM condition and 1200s in the Control condition. This was selected to provide sufficient time to complete approximately 60% of tasks, meaning that there was an advantage for allocating more time to tasks from the easy folder than from the hard folder. Four pilot participants completed 30 tasks in each condition. Mean task completion time was 2215s (SD = 215s) in the IRM condition and 1215s (SD = 235s) in the Control condition. These data were used to set the time limits for each condition such that the extra procedures in the IRM condition were accounted for and there was approximately the same amount of time to complete the actual tasks.

2.2. Results

In all three experiments reported, all analyses were initially conducted including the order the condition was completed (first or second) as a within participants variable. There were no interactions between condition order and any other independent variable analysed. Thus, to simplify the presentation of results, all tables, figures and statistics were computed from data aggregated across condition order. In all experiments, means were calculated per participant and analyses were conducted on these data. Reported means and standard deviations were calculated from the individual participant means. In all experiments, skewness values were between -2.56 and 2.56. All times are given in seconds.

In this experiment paired t tests were used to compare two means and, except when stated otherwise, 2 (Condition: IRM and Control) × 2 (Folder: Easy and Hard) ANOVAs with repeated measures on both factors were used for all other analyses. Tasks were rated as correctly completed when all the text had been entered, typing errors were ignored. There were no instances of three or more consecutive words being incorrectly entered.

2.2.1. Tasks Completed

Table 1 provides descriptive statistics for the completion of tasks in both easy and hard folders. The table shows that more tasks were completed from the easy folder than from the hard folder in both conditions, $F(1, 23) = 27.87$, $p < .001$, $\eta^2_p = .55$. There was no difference in mean number of tasks completed across both conditions and this did not interact with folder type. This pattern of effects was repeated for the total number of documents opened in both conditions (i.e. including instances where a document was opened but the task was not completed). More documents were opened from the easy folder than the hard folder, $F(1, 23) = 20.45$, $p < .001$, $\eta^2_p = .47$, and the difference between conditions and interaction were not significant.

The mean number of switches between easy and hard folders refers to the number of times that documents from different folders were opened consecutively. Table 1 shows there was approximately one switch between folders for every three tasks completed. The number of switches in the IRM and control conditions did not differ.

To understand which tasks were completed from the easy and hard folders, Figure 3 gives the proportion of participants that completed each task. The figure shows that the shortest 10 tasks from each folder were completed by 22 out of 24 participants. However, in the hard folder the proportion of participants that completed the remaining 15 tasks fell very steeply and only 1 participant completed the 9 longest tasks. In the easy folder the proportion of participants also dropped but not as steeply and only the longest 3 tasks were not completed by any participants. This pattern closely corresponds to the level of task difficulty as indexed by task length in number of characters shown in Figure 1. Thus, the task selection decisions enabled participants to complete the shortest tasks across both folders rather than spend time completing the longer tasks. These effects were the same in the IRM and control conditions.

2.2.2. Task Selection Times

Table 2 contains the mean time taken to select a task across all tasks in each condition. Task selection times were longer in the IRM condition than in the control condition, $t(23) = 6.85$, $p < .001$, $d = 1.15$. Task selection time was also computed excluding intervals where task performance was interrupted by access control procedures. This measure of uninterrupted

mean selection time in the IRM condition is given in Table 2 and was reliably longer than mean selection time in the control condition, $t(23) = 4.78$, $p < .001$, $d = .80$.

The number of between task intervals when both folders were opened is also given in Table 2. Participants checked both folders during more task intervals in the IRM condition than in the control condition, $t(23) = 3.27$, $p < .01$, $d = .72$. This also indicates task management processes were affected by the IRM-related interruptions. To test whether time spent checking other folders explained the differences in selection times, mean selection time was calculated excluding any intervals when both folders were checked (for the IRM condition intervals where email was checked or documents could not be opened were also excluded). Mean selection time was still reliably longer in the IRM condition than in the control condition, $t(23) = 4.57$, $p < .001$, $d = .77$.

2.2.3. Task management strategies

Table 1 shows that participants switched frequently between the folders and overall managed to select the shortest tasks to complete. Within each folder, selecting the shortest task was relatively simple as participants were informed that tasks were ordered from shortest at the top of the folder to longest at the bottom. Correlations were computed for each folder between the order tasks were completed and the position of the task in the folder. These indicated that participants simply worked from top to bottom for both folders in both conditions, IRM Easy, $r(22) = .99$; IRM Hard, $r(19) = .99$; Control Easy, $r(22) = .99$; Control Hard, $r(15) = .99$; All ps $< .001$.

Selecting the shortest task across folders was less straightforward, although comparison of Figures 1 and 3 indicates that participants achieved this. One strategy is to maximise the marginal rate of return or always complete the shortest available task. Because many of the tasks in each folder were of similar difficulty this strategy required participants to switch frequently between folders. We tested participants' strategies by comparing the marginal task completion time with the completion times for previous and subsequent tasks. The means and standard deviations are given in Table 3, and are categorised according to whether the next task completed was from the other folder (switch tasks) or from the same folder (no-switch

tasks). The means were analysed using a 2 (Condition: IRM and Control) × 2 (Task Type: Switch and No-Switch) ANOVA with repeated measures on both factors.

In general, mean task completion times were longer for tasks that participants switched after than for tasks they did not switch after, $F(1, 23) = 11.86$, $p < .01$, $\eta^2_p = .34$, thus supporting the theory that participants maximise the marginal rate of return. A more direct test is to compare the marginal task completion time in each folder after each task. Table 3 shows that when participants switched folders marginal task completion time was higher for the current folder than for the alternative folder, but when they did not switch the time was higher for the alternative folder than for the current folder. This difference between switch and no-switch decisions was significant, $F(1, 22) = 49.22$, $p < .001$, $\eta^2_p = .69$.

Unlike the participants, we have access to the actual task completion times for the next task in the alternative folder at each decision point. Thus, the current task completion time was compared with the task completion time for the next task in the alternative folder. This also found that when participants switched, the next task in the alternative folder was shorter than the current task but when they did not switch the next task in the alternative folder was longer than the current task. This difference between switch and no switch tasks was significant, $F(1, 23) = 31.93$, $p < .001$, $\eta^2_p = .58$.

Unfortunately, these last two comparisons are confounded in that, relative to corresponding items in the alternative folders, a switch task will always be completed after a no switch task. This means the switch task will typically be lower in the folder, and therefore longer, than the no switch task. These effects are still meaningful in that they reflect the actual decisions made by participants. Nonetheless, to correct for this confound, Table 3 also includes the difference between the current task and the next task irrespective of folder. This found that when participants switched task the current task was shorter than the next task and when participants did not switch task the current task was longer than the next task. Again the difference between switch and no switch tasks was significant, $F(1, 23) = 19.32$, $p < .001$, $\eta^2_p = .46$.

Therefore, all of these analyses indicate that participants switched task after longer tasks and that this switching improved the marginal rate of task completion. In all of the analyses on the data in Table 3 there were no differences between the IRM and Control conditions and no interactions.

2.2.4. Errors in Information Rights Management

Errors in applying the information rights management procedures are given as means per participant in Table 4. The types of errors were: an uncorrected failure to restrict access rights to a document (there were no incomplete procedures so all of these errors were caused by forgetting to complete the 5 step procedure), a corrected failure to restrict access rights (these documents were reopened and access rights were then restricted), a failure to enter the correct email address on the fifth step of the procedure and an error entering an email when requesting access to a document. An informal debrief indicated that, typically, participants were unaware of any errors made and thus would be unlikely to correct them.

2.3. Discussion

In both the IRM condition and the Control condition participants switched frequently between folders throughout the experiment and completed more tasks from the easy folder than from the hard folder. Thus, participants successfully managed their time to focus on the tasks that could be completed most quickly. The data indicate that they achieved this by maximising the marginal rate of return. That is, the frequent interleaving between folders was designed to ensure participants always worked on the shortest task available. These results are consistent with Payne et al. (2007) and generalise their findings to a more complex task environment.

The results provide an explanation for the frequent task switching observed in modern offices (Gonzalez and Mark, 2004) and also suggest the criterion that workers use to decide when to interleave between tasks. Furthermore, these findings are not inevitable. Given that participants knew the time available, there are a number of plausible strategies that would

have allowed them to complete more tasks from the easy folder than the hard folder with fewer switches between folders. We return to this issue in the General Discussion.

The other key finding from this experiment was that the IRM-related interruptions affected the processing involved in task management. In the IRM condition it took longer to select the next task to complete and participants were more likely to check both folders before choosing a task than in the Control condition. However, although task selection decisions took longer there were no differences between conditions in the actual decisions themselves.

Relative to the number of tasks completed the proportion of errors was relatively small, although even a single error can be costly for security and most participants were unaware of the errors that they had made.

3. Experiment 2

In Experiment 1 the documents within each folder were organised according to the length of the task. This mimicked the diminishing marginal returns over time within the scrabble tasks used in Payne et al. (2007). Interleaving between these tasks is interesting; not least because many everyday tasks contain diminishing marginal returns (i.e. consider the task of editing a document – over time fewer and fewer corrections are made). Nonetheless, we are also interested in task management when there are no diminishing marginal returns. Where tasks are arranged randomly within a folder users do not know how difficult the next task in the other folder will be. Therefore, we predict that users will no longer interleave frequently between folders as this will not necessarily increase the marginal rate of return.

Some clue about how users manage tasks in this situation can be found in work on skim reading. There users must decide whether reading the current section of text is more useful than reading some other, as yet unread, section of text. Duggan and Payne (2009, 2011) found skim readers satisficed, that is, they kept reading as long as the marginal rate of return

was above some threshold and then switched to the next section of text when the threshold was crossed. In this experiment, we randomised the order of tasks within each folder and predict that users will satisfice by opening each task in turn and only completing shorter tasks that have a sufficiently high marginal rate of return. Following this strategy would mean that more tasks will be completed from the easy folder because it contains more short tasks.

The access control manipulation from Experiment 1 was simplified to understand the effect on task selection times. In the IRM condition in this experiment, participants received four different emails at set junctures in the experiment, each requiring them to set the access rights to a different document. There was no longer a requirement to send emails requesting access to particular documents, nor were participants required to set access rights for any other documents. This design created four discrete interruptions that could be compared with the equivalent stage in the Control condition, providing a more controlled test of the effect of access control tasks upon overall task management processing. In addition, the interface was changed in both conditions such that after opening a document it no longer changed colour. This increased the difficulty of the task management as users had to rely on their memory. To increase the realism and complexity of the task environment distracter emails that participants did not have to respond to were included in the IRM condition.

The hypotheses are:

H_4: Participants will complete more tasks from the easy folder than the hard folder.

H_5: Task selection time will be longer in the IRM condition than the Control condition.

H_6: Task selection time following an interruption in the IRM condition will be longer than the matched juncture in the Control condition.

H_7: Following an interruption in the IRM condition more documents will be opened prior to completing a task than at the matched juncture in the Control condition.

3.1. Method

3.1.1 Participants

Participants were 10 students and 14 staff from the same population as used in Experiment 1, no participants had taken in part in Experiment 1. They received £10 for taking part in the experiment. Thirteen participants were female and the age range was 18-47 years (M = 27.92, SD = 8.38). All participants reported they were familiar with Microsoft Office and had used it for more than 10 hours.

3.1.2. Materials

The documents and tasks from Experiment 1 were presented using the same purpose-written program with the following modifications. Within each folder documents were positioned in a random order from top to bottom and no documents were labelled in red or treated as priority tasks. When a document or task had been opened, the label within the folder did not change to blue and remained the same colour. No other information that participants could use to aid their memory was provided within the interface (e.g. a timestamp of the last time the document was edited) nor were participants allowed to make notes during the task.

For the IRM condition there were no documents for which access was denied. For each pair of folders 12 new email messages were created (meaning 24 new messages and 2 new practice messages in total). Four of each set of 12 messages gave specific instructions to allocate access rights to a particular document to a specified email address. The remaining eight emails were distracter items that did not refer to the experiment (i.e. they referred to hypothetical meetings, contained spam etc.). In the IRM condition emails "arrived" within the email package at fixed times, these time junctures in seconds were: 100, 210, *340*, 470, *480*, 610, 790, 800, *810*, 1010, *1120*, 1340 (italics indicate an interruption i.e. an email that required participants to grant access rights).

When completing the access control procedure the email address was not displayed within the word simulation, instead participants had to refer back to the appropriate message in the email package.

3.1.3. Design

As in Experiment 1, in the IRM condition any time spent using the email package or adding access rights to documents was subtracted from the task selection time. The measure of uninterrupted selection time excluded all intervals where participants added access rights to a document. Where participants checked email without adding access rights, the time spent checking mail was excluded but the remainder of the interval was included in the mean task selection time.

The end of the interruption was deemed to be the first time they opened a folder after completing all email interactions and access control procedures (including reopening documents they granted access rights to). At the four interruption points, two measures of task selection time are reported: "Selection time for next document opened" was measured from the end of the interruption until any document was opened. "Selection time for next task completed" was measured from the end of the interruption until a document was opened in which the associated task was completed (i.e. including time spent selecting and opening documents for which the task was not completed).

For each of the 4 interruptions in the IRM condition a matched juncture in the Control condition was calculated as being at the next interval between tasks after the same proportion of total task time had elapsed.

3.1.4. Procedure

Participants received the same three practice tasks as in Experiment 1 but access was not denied to any of the documents. One email during practice required participants to grant access rights to a particular document and a distracter email was also included.

The written instructions were modified to explain that documents were presented in a random order within each folder. In the IRM condition participants were instructed that emails would arrive throughout the experiment. They were told to grant access to documents immediately when requested to do so and to ignore all other emails. Unlike in Experiment 1 participants were not asked to add access rights to a document after completing a task. There were no emails and no access to the email package in the Control condition.

Total time for task completion was 1500s in the IRM condition and 1200s in the Control condition. Piloting of the materials indicated that this difference accounted for the extra procedures in the IRM condition and left approximately the same amount of time to complete the actual tasks.

All other aspects of the method were the same as in Experiment 1.

3.2. Results

As in Experiment 1 paired *t* tests were used to compare two means and, except when stated otherwise, 2 (Condition: IRM and Control) × 2 (Folder: Easy and Hard) ANOVAs with repeated measures on both factors were used for all other analyses. Tasks were rated as correctly completed when all the text had been entered, typing errors were ignored. There were no instances of three or more consecutive words being incorrectly entered.

3.2.1. Tasks Completed

From each folder the mean number of tasks that were completed, the mean number of tasks that were opened and the mean number of switches between folders are given in Table 5. More tasks were completed from the easy folder than the hard folder in both conditions, $F(1, 23) = 79.49$, $p < .001$, $\eta^2_p = .78$. There was no difference in the number of documents that were opened from each folder. For both measures there was no difference between the IRM and control condition nor were there any interactions. The mean number of switches between folders was much lower than in Experiment 1. There was no difference in number of switches between the IRM and control conditions.

The proportion of participants that completed each task is given in Figure 4. The figure shows that the majority of participants completed the shortest 10 tasks in both folders. However, the majority of participants also completed all but the 3 longest tasks in the easy folder, whereas the proportion of participants who completed the 15 longest tasks in the hard folder was

much lower. As in Experiment 1, the proportion of participants who completed each task closely corresponded with the level of task difficulty as indexed by task length in number of characters shown in Figure 1. These patterns were the same in both the IRM and control conditions.

3.2.2. Task Selection Times

Table 6 contains mean task selection times across the task and directly after interruptions. In the IRM condition, task selection took more time than in the control condition, $t(23) = 5.09$, $p < .001$, $d = 1.09$. Excluding intervals containing interruptions to add access rights did not change the pattern of this effect and uninterrupted IRM task selection times were longer than the control task selection times, $t(23) = 2.17$, $p < .05$, $d = .45$. Table 6 also shows there were more intervals where both folders were checked in the IRM condition than in the control condition, $t(23) = 2.24$, $p < .05$, $d = .52$. The intervals where both folders were checked and intervals where access control procedures were completed were then excluded from the task selection times. This adjusted measure also indicated that task selection was longer in the IRM condition than in the control condition, $t(23) = 2.08$, $p < .05$, $d = .37$.

The mean task selection time following an interruption to set access rights is given for the IRM condition in Table 7. Also included is the mean task selection time for the matched junctures in the control condition. The means indicate that immediately after an interruption in the IRM condition participants took longer to select the next task than at the same point in the control condition, $t(23) = 4.60$, $p < .001$, $d = 1.04$. In the control condition, the mean of the task selection times matched against interruptions was longer than the overall mean task selection time. One explanation for this is that the overall mean task selection time included a number of very short selections towards the end of the experiment. Also given in Table 7 is the time to select the next task that was completed. This time included any instances where a document was opened but the task was not completed. This selection time was longer post-interruption in the IRM condition than in the control condition, $t(23) = 3.68$, $p < .01$, $d = .72$.

The final measures in Table 7 compare the number of documents that were opened or reopened immediately after an interruption in the IRM condition and at the same point in the control condition. The means show that before selecting a task that was completed more

documents were opened, $t(23) = 2.44$, $p < .05$, $d = .46$, and more documents were reopened, $t(23) = 3.20$, $p < .01$, $d = .82$, in the IRM condition than in the control condition.

3.2.3. Task Management Strategies

The task selection decision was very different to Experiment 1 because tasks were ordered randomly within folders, therefore within a folder the only way a participant could find out the task difficulty was by opening the document. This was reflected in the high proportion of documents that were opened and then closed without completing the task. Further, frequent switching between folders would not support maximisation of the marginal rate of return and, as predicted, there were few switches between folders. Nonetheless, Figure 4 indicates the participants successfully allocated their time such that they completed the shorter tasks and did not complete the longer tasks.

Given that a high proportion of documents were opened without completing the task, strategy was analysed using documents opened rather than tasks completed. Correlations between position of task within a folder and the order that the first 25 documents from each folder were opened were high for both easy and hard folders in both conditions, IRM Easy, $r(25) = .99$; IRM Hard, $r(25) = .96$; Control Easy, $r(25) = .99$; Control Hard, $r(25) = .99$; All ps $< .001$. Thus, participants worked from top to bottom opening documents consecutively within each folder and closed the documents containing longer tasks without completing them (because only one document could be open at a time each document was "closed" by opening another document). The effect of an interruption upon this strategy was given by the proportion of times that participants opened the next document down within a folder. Following an interruption in the IRM condition, the mean proportion of times participants opened the next document was .30 ($SD = .28$), whereas at the matched juncture in the control condition the mean proportion was .89 ($SD = .21$). This difference was significant, $t(23) = 8.45$, $p < .001$, $d = 1.53$.

3.2.4. Errors in Information Rights Management

Across all participants there were 4 failures to grant restricted access to a document (.04 proportion of all requests). Participants did not initiate the IRM procedure in any of these

instances indicating this reflected forgetting rather than an inability to complete the procedure.

3.3. Discussion

As in Experiment 1, participants completed more tasks from the easy folder than the hard folder. However, unlike Experiment 1 there were very few switches between folders and there was no difference between the number of documents opened in either folder. Therefore, participants managed to adaptively allocate more time to the shorter tasks without interleaving between folders. Investigation of the task management strategies indicated they achieved this by satisficing – opening each document in turn and only completing the shorter tasks that had a sufficiently high marginal rate of return. The low rate of interleaving between folders indicates that participants will not switch frequently between tasks when this does not increase the marginal rate of return.

Task selection times were longer in the IRM condition than in the Control condition. This replicates and extends this finding from Experiment 1, as the access control tasks were fewer and less complex than in Experiment 1. Moreover, immediately after an access control task participants were more likely to (1) open a task they did not complete, (2) reopen a task that had already been completed and (3) open a non-consecutive task. Thus, task management processing was slower and more likely to lead to errors following an access control task. The final experiment investigated the consequences of errors in task management for overall task performance.

4. Experiment 3

Experiments 1 and 2 demonstrated that participants are able to manage their time across tasks to complete shorter tasks when their goal was to maximise the total number of tasks completed. Their strategies to achieve this were consistent with maximising the marginal rate of return. Both experiments also showed that interruptions from access control tasks affect task management processing and can lead to errors in task management. Therefore,

interleaving between tasks can improve performance by focusing on the task that produces the greatest short-term gain but can also interfere with performance by interrupting task management. This experiment investigated both of these effects by alerting the participant when an email arrived.

Russell et al. (2007) found that most users received alerts when email arrived and that this system suited them. Alerts support task management by increasing users' awareness of other tasks, but because they are an interruption, they can disrupt primary task performance (Bailey and Konstan, 2006) and Experiments 1 and 2 indicate that they could also disrupt task management. To test the effect of alerts, in this experiment all conditions received access control tasks by email. However, only half of the participants received alerts for these emails and the remainder had to check their email spontaneously as in the previous experiments.

Experiments 1 and 2 demonstrated that interleaving strategies and interruptions affect task selection, however, failures of task management could also affect primary task performance. To test this, the design and hypotheses in Experiment 3 focused on task performance. This meant more cognitively demanding tasks were used, in some conditions participants had to send emails to request access to some documents and there was an incentive to complete all the tasks as quickly as possible. These changes increased the complexity of the task environment and provided a more powerful test of the effect of interruptions upon task performance.

Following Bailey and Konstan (2006), we predict that the alerts will disrupt performance on the primary task. However, Russell et al. (2007) indicated that alerts can increase user awareness, therefore, we predict that when there are no alerts participants will be more likely to forget to complete the access control tasks. In addition to manipulating alerts we also varied the amount of interleaving required by participants by increasing the number of access control tasks. We predict that increasing the interleaving will increase the cognitive load and reduce the available mental resources and thus amplify both the costs to primary task performance and the benefits to access control performance provided by alerts.

The hypotheses are:

H_8: More primary task errors will be made in the Alert condition than in the No Alert condition.

H_9: More access control errors will be made in the No Alert condition than in the Alert condition.

H_{10}: In the High interleaving condition, alerts will impair primary task performance more than in the Low interleaving condition.

H_{11}: In the Low interleaving condition, alerts will impair access control task performance more than in the High interleaving condition.

4.1. Method

4.1.1. Participants

Participants were 32 students from the same population as used in Experiments 1 and 2, no participants had taken in part in either Experiment 1 or 2. Participants did not receive any financial compensation for taking part. Eleven participants were female and the majority (80%) of participants were aged between 21-25. All participants reported they were familiar with Microsoft Office and had used it for more than 10 hours.

4.1.2. Materials

Two sets of eight comprehension tasks were created and each task was divided into three subtasks. Each subtask comprised a document and a question about information within the document. For example, a document containing record temperatures in different countries was paired with the question "What is the highest temperature ever recorded in Europe?". The documents were different from those used in Experiments 1 and 2 and were selected from a range of different websites. Questions were chosen such that information had to be read carefully to avoid mistakes. Questions were pretested using participants from the same population who did not take part in any of the 3 experiments. Pretesting revealed that each question was answered correctly by the majority of participants but that errors were made due

to participants failing to read the document carefully. The three subtasks that made up a task referred to a similar topic (i.e. politics, health etc.).

The program from Experiment 2 was modified to contain only one folder. This folder contained all 24 documents for a set of eight tasks. Each document was labelled using the topic name and either 1, 2 or 3. The documents were organised by name and then by number and only one document could be opened at a time. The first document in a task included all three questions for that task with textboxes for the answers and also the first document. The other two documents within a task just contained the documents. When a document or task had been opened, the label within the folder did not change to blue and remained the same colour. There were no priority tasks labelled in red.

In the High Interleaving condition, access was denied to 6 of the 24 documents, clicking on these documents produced the same dialog box as in Experiment 1. As in Experiment 2, four email messages were received that required participants to enable particular email addresses to access particular documents. The first message arrived after 100 seconds and there was a gap of 150-200 seconds between each of the three subsequent messages. Three distracter email messages similar in format to those from Experiment 2 were sent after 100, 500 and 900 seconds.

Alerts for email messages were a grey box that appeared in the bottom right hand corner of the screen for four seconds instructing the participant that they had new email and including the name of the sender.

All other features of the program were the same as in Experiments 1 and 2.

4.1.3. Design

The presence or absence of email alerts (Alert vs. No Alert) was manipulated between participants and the amount of interleaving required (High Interleaving vs. Low Interleaving) was manipulated within participants in a 2 × 2 mixed design. The set of comprehension tasks

received first and the set of comprehension tasks allocated to High and Low Interleaving conditions were counterbalanced across participants.

The stage in task completion at which participants checked email was used to assess interleaving strategy. If participants opened the email package after entering an answer to a question and before opening another document it was categorised as being between subtasks. All other instances in which participants opened the email were categorised as being in the middle of a subtask.

4.1.4. Procedure

Participants were first provided with written instructions and then completed a practice task containing three subtasks. Access to one document from the practice task was restricted and participants were required to send an email requesting access and wait for the reply before being able to access the document. At the end of the practice task the participant was required to set access rights for the first document such that a specified email address was given access.

Participants were divided into two equal sized groups. One group received an alert for every incoming email, the other group did not and had to check the email package to see if any email had arrived. In every condition, participants were instructed to complete all 24 comprehension subtasks in any order they chose and that answers to the questions should be entered in the textbox provided in the first document for each task. They were also told to complete any access control tasks sent by email. It was emphasised that emails did not have to be responded to immediately but neither should they be left until the end of the experiment and completed then.

All conditions received four emails during the experiment that required them to set access rights to particular documents. The procedure for setting access rights was the same as in Experiments 1 and 2. As in Experiment 2, participants were not required to set access privileges for any other documents. The only difference between the High and Low Interleaving conditions was that access was denied to some documents in the High

Interleaving condition. The procedure for gaining access was the same as in Experiment 1, replies to emails that requested access arrived between 30-60 seconds after the message had been sent.

Participants were told that the three fastest times to complete all the comprehension subtasks and access control tasks would receive a prize. They were also given a nominal time limit of 2400 seconds to complete all the tasks in each condition. All participants in all conditions reported they had completed the tasks before this time limit had elapsed. After completing a set of eight comprehension tasks in either the High or Low Interleaving condition they then completed the other set of eight comprehension tasks in the other condition.

4.2. Results

The proportion of comprehension subtasks correctly completed did not differ across the 2 sets of tasks (.87 and .88) and task set did not interact with any other variables in any of the analyses. Thus, to simplify the presentation of results, all tables and statistics were computed from data aggregated across task set. Results were analysed using 2 (Email notification: Alert vs. No Alert) × 2 (Interleaving: High vs. Low) mixed ANOVAs with repeated measures on the second variable.

4.2.1. Task Performance

Descriptive statistics for performance on the comprehension tasks and on restricting access to documents in each of the conditions are given in Table 8. The time to complete all 24 comprehension subtasks and any access control tasks was significantly longer in the High Interleaving conditions than the Low Interleaving conditions, $F(1, 30) = 62.60$, $p < .001$, $\eta^2_p = .68$. This was unsurprising given there were a number of extra procedures that had to be completed in the High Interleaving condition. There were no other significant differences or interactions for time to complete the task.

Table 8 shows that a higher proportion of comprehension subtasks were correctly completed in the No Alert condition relative to the Alert condition, $F(1, 30) = 4.83$, $p < .05$, $\eta^2_p = .14$ Table 8 also shows that the proportion of access control tasks completed was higher in the Alert condition than the No alert condition, $F(1, 30) = 4.66$, $p < .05$, $\eta^2_p = .13$. Therefore, providing alerts degraded comprehension task performance but improved performance on the access control tasks. There were no other reliable effects or interactions upon proportion of comprehension subtasks completed or proportion of access control tasks completed.

Participants did not provide an answer for 20 of the 195 incorrect comprehension subtasks indicating these tasks were forgotten rather than wrongly answered. For all the access control tasks not completed, the IRM procedure was not attempted. This indicated the access control task errors were due to forgetting rather than an inability to complete the procedure.

4.2.2. Task Interleaving

To understand the effects of email alerts upon task performance we analysed the interleaving strategy in each of the four conditions. Table 9 gives the mean number of times participants interleaved between a document and the email package in the middle of a comprehension subtask or between subtasks. Participants were more likely to switch to email in the middle of a subtask in the Alert condition than in the No Alert condition, $F(1, 30) = 9.02$, $p < .01$, $\eta^2_p = .23$. Between subtasks there was no difference in the Alert and No Alert conditions in the frequency of interleaving between email and the documents. There was more interleaving in the high interleaving condition than the low interleaving condition in the middle of a subtask, $F(1, 30) = 152.09$, $p < .001$, $\eta^2_p = .84$, and between subtasks, $F(1, 30) = 13.83$, $p < .01$, $\eta^2_p = .32$. There were no significant interactions.

4.3. Discussion

More comprehension subtasks were correctly completed in the No Alert condition than in the Alert condition indicating that receiving alerts degraded primary task performance. In contrast, more access control tasks were completed in the Alert condition than in the No Alert condition. The interleaving strategy also differed in that participants who received alerts were

more likely to switch away from the comprehension task in the middle of a subtask. Taken in combination these results suggest that, on receiving an email alert, participants interrupted the comprehension task to complete the access control task. This helped them to remember to complete the access control task but led to errors on the primary task. There was in effect a trade-off between performance on the comprehension task and the access control task.

This trade-off demonstrates that interleaving can both improve task performance (more access control tasks completed in the Alert condition) and degrade task performance (fewer comprehension subtasks completed in the Alert condition). This underscores the importance of understanding when participants interleave between tasks and demonstrates that simple alterations to software (i.e. providing alerts) can have important consequences for task performance.

As in Experiments 1 and 2, the decisions to switch task in this experiment can be interpreted as maximising the marginal rate of return. When participants received an alert, they could be certain that an email had arrived, this increased the value of the alternative task (checking email) and increased the marginal rate of return for switching task. In the No Alert condition, it was uncertain whether there would be new email, therefore, checking email typically had a lower marginal rate of return than after an Alert. This meant that in the No Alert condition, participants were more likely to wait until the end of the comprehension subtask when the marginal rate of return for the comprehension task was lower. One explanation for this lower marginal rate of return could be lower workload at subtask boundaries (Bailey and Iqbal, 2008). We shall discuss the applicability of these ideas further in the General Discussion.

Counter to our hypotheses, the amount of interleaving did not affect the extent to which alerts disrupted performance as there was no interaction between High/Low Interleaving and Alert/No Alert. This may be because there was still a large amount of interleaving required in the Low Interleaving condition. Either way, the presence of both main effects means that the absence of an interaction does not challenge our theoretical account.

5. General Discussion

Across three experiments we have demonstrated that users will interleave between tasks to focus on shorter tasks and that these switching decisions are consistent with a strategy of maximising the marginal rate of return. The experiments also show that interruptions from access control tasks disrupt the processing involved in task management and the timing of task switches.

In Experiments 1 and 2 participants successfully allocated more time to a folder containing shorter tasks, thus increasing the number of tasks completed. In Experiment 1 participants achieved this by switching between folders to complete the shortest task available. In Experiment 2 participants achieved this by abandoning tasks that took longer to complete. Both of these interleaving strategies enabled participants to maximise the marginal rate of return. In Experiment 3, providing alerts for incoming emails changed the marginal rate of return and led to participants checking email in the middle of a subtask.

Experiments 1 and 2 also found that adding access control to documents increased the time to select the next task. Moreover, in Experiment 2 task management was disrupted such that participants made more errors in task selection following an access control task. Experiment 3 applied this finding to show that email alerts helped participants to remember access control tasks but increased the number of errors on the comprehension task.

These results lead to three key conclusions. First, task interleaving can help performance by enabling users to work on the most productive task. Second, users spontaneously adopt strategies that support performance by maximising the marginal rate of return. Third, task management can lead to errors that affect task performance. Next, we provide an account of when participants choose to interleave between tasks.

5.1. Maximising the marginal rate of return

Following Pirolli (2007) we have drawn on the animal foraging literature to understand human behaviour. Thus, like Payne et al. (2007), we have described interleaving behaviour in terms of rates of return. That is, task performance can be understood as the accumulation of some task currency, meaning that decisions to switch between tasks are based on the rate at which the currency is being accumulated on each of the tasks. We argue that the interleaving in our experiments enabled participants to maximise the marginal rate of return. That is, participants selected the task that provided the greatest benefit at that moment. Furthermore, we suggest that the frequent switching observed in modern workplaces (Czerwinski et al., 2004; Gonzalez and Mark, 2004) reflects a tendency to maximise the marginal rate of return.

We also argue that maximising the marginal rate of return is not the only available strategy or necessarily the most effective strategy. For example, in Experiment 1 participants could have switched between folders fewer times yet spent the same proportion of time on tasks from the easier folder. Or they could have simply spent half the time on each folder and switched once – a control condition in Payne et al. (2007) found benefits from working on the same task for longer periods of time. Similarly in Experiment 2, rather than satisficing – opening each document in turn and only completing the shorter tasks that had a sufficiently high marginal rate of return – participants could have "sampled" (Reader and Payne, 2007). That is, prior to completing tasks, they could have opened several documents from each folder to compare the difficulty of the folders. These and other strategies indicate that it was not inevitable that our participants appeared to maximise the marginal rate of return.

Where there are diminishing marginal returns or where the task environment changes rapidly, maximising the marginal rate of return will lead to frequent task switching. However, frequent interleaving between tasks can lead to errors in task performance. In Experiment 3, providing email alerts encouraged participants to switch in the middle of a subtask, which led to task completion errors. Therefore, we suggest that in some environments frequent interleaving by users will impair their overall task performance (Iqbal and Horvitz, 2007; Mark et al., 2012). Clearly, the most appropriate interleaving strategy will vary according to the environment, the task and the user's knowledge. However, our experiments indicate that the perceived marginal rate of return will predict when users will interleave between tasks.

5.2. Future theoretical extensions

Charnov's (1976) marginal value theorem describes animal foraging behaviour at a general level and subsequent research has specified strategies used by individual animals to maximise the marginal value (e.g. Waage, 1979; Iwasa et al., 1981). Analogously, further work is needed to identify task interleaving strategies adopted by users to maximise the marginal rate of return. Satisficing is one such potential strategy and we suggest it could explain the interleaving behaviour in Experiment 2. A more detailed computational model was developed by Payne et al. (2007) to explain the interleaving behaviour on their tasks. In short, they claimed that a combination of Green's rule (1984) and switching at subgoals explained their data. Further research needs to test these and other explanations and identify the conditions under which they apply.

In Experiment 3, participants often checked email shortly after receiving an alert. However, it is unclear whether this reflected their everyday behaviour upon receiving an alert or was a demand characteristic of the experiment. This example highlights the general difficulty of establishing exactly which currency participants are maximising. They may be seeking to complete the experiment quickly, accurately or with as little effort as possible. This makes it difficult to test predictions about the marginal rate of return. Future experiments could lessen this problem by explicitly providing a currency (i.e. points) within the experiment that users must maximise. This would enable us to draw stronger conclusions but could also make the interleaving behaviour less representative of everyday task completion.

Specifying a task currency also enables inferences about the rationality of users' behaviour. Unlike Charnov (1976) our account does not state that task (patch) switches are optimal. Rather, we suggest that users select the task with the highest *perceived* marginal value. This approach draws on Howes et al.'s (2009) use of subjective expected utility to assess the rationality of users. The value for this approach in understanding the rationality of interleaving decisions has been shown through experimental work by Janssen and Brumby (2010; Janssen et al., 2011).

5.3. Application of Findings

The primary tasks used in all three experiments were representative of typical everyday tasks completed in office environments (i.e. adding text to documents, answering questions by retrieving information from documents). The access control tasks used a widely available information rights management system that, due to its novelty, participants were not proficient in but may become widely adopted. The operating system, word processing and email software were familiar to all the participants. These factors were designed to increase the ecological validity of the findings within a controlled laboratory environment that enabled us to test specific predictions about interleaving strategies. Conducting these experiments in the field would have removed factors particular to the laboratory but also introduced a range of factors specific to that particular field environment that may not generalise to other field environments. Furthermore, much of the email and interleaving behaviour observed was consistent with that reported from field studies (Gonzalez and Mark, 2004; Jin and Dabbish, 2009; Russell et al., 2007). Notwithstanding this, a natural next step is to extend our methods to an actual office environment.

Access control tasks disrupted task management in all three experiments. Thus, where there are interruptions, task management is more likely to lead to errors. This effect was shown clearly in Experiment 3 where differences in interleaving strategy directly affected the number of errors. The straightforward implication from these results is that software designers need to be aware that users are likely to make errors when interleaving rapidly between tasks. Further, results found that these errors can be of more than one type – forgetting to complete a particular task or making an error during a particular task. Experiment 3 also demonstrates that alerts are a specific form of this problem, with users switching task in the middle of a subgoal rather than between subgoals.

Bailey and Iqbal (2008) have demonstrated that workload is lower upon completion of subgoals and that subtask boundaries are therefore more appropriate junctures for interruptions. Following on from this, Iqbal and Bailey (2010) describe a system that supports task management by providing alerts at subtask boundaries. Our experiments reinforce these findings and, by requiring users to complete multiple tasks and providing choice over when

to interleave between tasks, we have demonstrated that users will interrupt subtasks upon receiving an alert and that the costs of an interruption apply to task management as well as performance on the interrupted task. Relatedly, Salvucci and Bogunovich (2010; Bogunovich and Salvucci, 2011) have shown that users will choose to attend to an interruption when workload is low but that this decision is also affected by the time constraints of the interruption.

Other related work

The trade-off in performance between the comprehension task and the access control task in Experiment 3 indicates the importance of user motivation. Where users have discretion, they will prioritise some tasks over others. Duggan et al. (2012) showed that many users have a low motivation to complete security tasks thus, where they have discretion, users are likely to switch to other tasks that have a higher marginal rate of return. This is a particular problem where a security breach can lead to large costs for the organisation but completing the security task produces a low marginal rate of return for the individual. Our findings emphasise that where users are frequently interrupted and/or security systems are interleaved with other tasks, organisations should account for the likelihood that security may be compromised.

Our experiments also showed that users are able to allocate their time adaptively across multiple tasks by interleaving between them. This implies that providing users with discretion over their task management can increase the amount of work completed. This is particularly relevant in dynamic environments where task interleaving can be most effective. As noted above further research could focus on the specific strategies employed from which more situation specific guidelines could be inferred (i.e. see Duggan & Payne, 2011 for specific guidelines relating to skim reading). More speculatively, the implications for maximising the marginal rate of return may be less positive in situations where the benefits for continued work on a task cannot be easily translated into a marginal rate of return (insight problems are an extreme example of this type of task). Traditionally switch costs have discouraged premature switching from tasks that have little immediate reward, however, technological developments have reduced these switch costs meaning that sustained work on the same task

is less likely. Following the claims of some commentators (e.g. Gasser and Palfrey, 2009; Rosen, 2008; Willingham, 2010) we believe this is another interesting area for future exploration.

Acknowledgements

This work was supported by grant P0007E ("Trust Economics") from the UK Technology Strategy Board (TSB). We would like to thank members of the Trust Economics team for their input.

References

Adamczyk, P. D., Bailey, B. P., 2004. If not now when? The effects of interruptions at different moments within task execution. *Proceedings of the ACM Conference on Human Factors in Computing Systems, CHI '04.* Vienna, Austria, 271–278.

Altmann, E. M., Trafton, J. G., 2002. Memory for goals: an activation-based model. *Cognitive Science, 26*, 39-83.

Bailey, B. P., Konstan, J. A., 2006. On the need for attention aware systems: Measuring effects of interruption on task performance, error rate, and affective state. *Computers in Human Behavior, 22*, 685-708.

Bailey, B. P., Iqbal, S. T., 2008. Understanding changes in mental workload during execution of goal-directed tasks and its application for interruption management. *ACM Transactions on Computer-Human Interaction. 14, 4*, 1–28.

Beautement, A., Coles, R., Griffin, J., Ioannidis, C., Monahan, B., Pym, D., Sasse, M. A., & Wonham, M., 2008. Modelling the human and technological costs and benefits of USB memory stick security. *Proceedings of the 8th Workshop on Economics and Information Security, WEIS '08*. Dartmouth, USA.

Benbunan-Fich, R., Adler, R. F., Mavlanova, T., 2011. Measuring multitasking behaviour with activity-based metrics. *ACM Transactions on Computer-Human Interaction, 18, 2*, Article 7.

Bogunovich, P., Salvucci, D. D., 2011. The effects of time constraints on user behaviour for deferrable interruptions. *Proceedings of the ACM Conference on Human Factors in Computing Systems, CHI '11*, Vancouver, Canada, 3123-3126.

Bradwell, P., Reeves, R., 2008. Network citizens: Power and responsibility at work. http://www.demos.co.uk/ publications/networkcitizens.

Brumby, D. P., Salvucci, D. D., Howes, A., 2009. Focus on driving: How cognitive constraints shape the adaptation of strategy when dialing while driving. *Proceedings of the ACM Conference on Human Factors in Computing Systems, CHI '09,* Boston, USA, 1629-1638.

Carminati, B., Ferrari, E., Perego, A., 2009. Enforcing access control in web-based social networks. *ACM Transactions on Information and System Security, 13*, Article 6.

Cades, D. M., Boehm-Davis, D. A., Trafton, J. G., Monk, C. A., 2011. Mitigating disruptive effects of interruptions through training: What needs to be practiced? *Journal of Experimental Psychology: Applied, 17, 2*, 97-109.

Cao, X., Iverson, L., 2006. Intentional access management: Making access control usable for end-users. *Symposium On Usable Privacy and Security, SOUPS '06*. Pittsburgh, USA.

Czerwinski, M., Horvitz, E., Wilhite, S., 2004. A diary study of task switching and interruptions. *Proceedings of the ACM Conference on Human Factors in Computing Systems, CHI '04.* Vienna, Austria, 175-182.

Charnov, E. L., 1976. Optimal foraging: The marginal value theorem. *Theoretical Population Biology, 9*, 129-136.

Duggan, G. B., Payne, S. J., 2009. Text skimming: The process and effectiveness of foraging through text under time pressure. *Journal of Experimental Psychology: Applied, 15*, 228-242.

Duggan, G. B., Payne, S. J., 2011. Skim reading by satisficing: Evidence from eye tracking. *Proceedings of the ACM Conference on Human Factors in Computing Systems, CHI '11*, Vancouver, Canada, 1141-1150.

Duggan, G. B., Johnson, H., Grawemeyer, B., 2012. Rational security: Modelling everyday password use. *International Journal of Human -Computer Studies, 70,* 415-431.

Gasser, U., Palfrey, J., 2009. Mastering multitasking. *Educational Leadership, March 2009*, 15-19.

Gillie, T. Broadbent, D., 1989. What makes interruptions disruptive? A study of length, similarity, and complexity. *Psychological Research. 50, 4*, 243–250.

Gonzalez, V., Mark, G., 2004. "Constant, constant, multitasking craziness": Managing multiple working spheres. *Proceedings of the ACM Conference on Human Factors in Computing Systems, CHI '04*. Vienna, Austria, 113-130.

Green, R. F., 1984. Stopping rules for optimal foragers. *American Naturalist, 123*, 30-43.

Herley, C., 2009. So long and no thanks for the externalities: The rational rejection of security advice by users. *Proceedings of the New Security Paradigms Workshop, NSPW '09*. Oxford, UK.

Hodgetts, H. M., Jones, D. M., 2006. Interruption of the Tower of London task: Support for a goal activation approach. *Journal of Experimental Psychology: General, 135*, 103 – 115.

Howes, A., Lewis, R. L., Vera, A., 2009. Rational adaptation under task and processing constraints: Implications for testing theories of cognition and action. *Psychological Review, 116*, 717-751.

Iqbal, S. T., Bailey, B. P., 2010. Oasis: A framework for linking notification delivery to the perceptual structure of goal-directed tasks. *ACM Transactions on Computer-Human Interaction, 17 (4),* Article 15.

Iqbal, S. T., Horvitz, E., 2007. Disruption and recovery of computer tasks: Field study, analysis and directions. *Proceedings of the ACM Conference on Human Factors in Computing Systems*; *CHI '07,* San Jose, USA, 677-686.

Iwasa, Y., Higashi, M., Yamamura, N., 1981. Prey distribution as a factor determining the choice of optimal foraging strategy. *American Naturalist 117*, 710-723.

Jackson, T. W., Dawson, R. J., Wilson, D., 2001. The cost of email interruption. *Journal of Systems and Information Technology, 5*, 81–92.

Janssen, C. P., Brumby, D. P., 2010. Strategic adaptation to performance objectives in a dual-task setting. *Cognitive Science, 34*, 1548-1560.

Janssen, C. P., Brumby, D. P., Dowell, J., Chater, N., Howes, A., 2011. Identifying optimum performance trade-offs using a cognitively bounded rational analysis model of discretionary task interleaving. *Topics in Cognitive Science, 3*, 123-139.

Jin, J., Dabbish, L. A. Self-interruption on the computer: A typology of discretionary task interleaving. *Proceedings of the ACM Conference on Human Factors in Computing Systems*; *CHI '09*, Boston, USA, 1799-1808.

Mangelsdorf, M. E., 2011. How too much multitasking at work can slow you down. MIT Sloan Management Review, 52, 3, 96.

Mark, G., Gonzalez, V. M., Harris, J., 2005. No task left behind? Examining the nature of fragmented work. *Proceedings of the ACM Conference on Human Factors in Computing Systems*; *CHI '05*, Portland, USA, 321-330.

Mark, G., Gudith, D., & Klocke, U., 2008. The cost of interrupted work: More speed and stress. *Proceedings of the ACM Conference on Human Factors in Computing Systems*; *CHI '08*, Florence, Italy, 107-110.

Mark, G., Voida, S., Cardello, A. V., 2012. "A pace not dictated by electrons": An empirical study of work without email. *Proceedings of the ACM Conference on Human Factors in Computing Systems*; *CHI '12*, Austin, USA, 555-564.

McFarlane, D. C., Latorella, K. A., 2002. The scope and importance of human interruption in human-computer interaction design. *Human-Computer Interaction, 17*, 1-61.

Payne, S. J., Duggan, G. B., Neth, H., 2007. Discretionary task interleaving: Heuristics for time allocation in cognitive foraging. *Journal of Experimental Psychology: General, 136*, 370-388.

Pirolli, P., 2007. *Information foraging theory: Adaptive interaction with information*. Oxford: Oxford University Press.

Reader, W. R., Payne, S. J., 2007. Allocating time across multiple texts: Sampling and satisficing. *Human-Computer Interaction, 22*, 263-298.

Rogers, R. D., Monsell, S., 1995. Costs of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology: General, 124*, 207-231.

Rosen, C., 2008. The myth of multitasking. *The New Atlantis, Spring 2008*, 105-110.

Rubinstein, J. S., Meyer, D. E., Evans, J. E., 2001. Executive control of cognitive processes in task switching. *Journal of Experimental Psychology: Human Perception and Performance, 27, 4*, 763-797.

Russell, E., Purvis, L. M., Banks, A., 2007. Describing the strategies used for dealing with email interruptions according to different situational parameters. *Computers in Human Behavior, 23*, 1820-1837.

Salvucci, D. D., Bogunovich, P., 2010. Multitasking and monotasking: The effects of mental workload on deferred task interruptions. *Proceedings of the ACM Conference on Human Factors in Computing Systems, CHI '10*, Atlanta, USA, 85-88.

Salvucci, D. D., Taatgen, N. A., 2008. Threaded cognition: An integrated theory of concurrent multitasking. *Psychological Review, 115*, 101-130.

Sandhu, R. S., Coyne, E. J., Feinstein, H. L., Youman, C.E., 1996. Role-based access control models. *IEEE Computer, 29*, 38-47.

Smetters, D. K., Good, N., 2009. How users use access control. *Symposium On Usable Privacy and Security, SOUPS '09.* Mountain View, USA.

Speier, C., Valacich, J. S., Vessey, I., 1999. The influence of task interruption on individual decision making: An information overload perspective. *Decision Sciences, 30*, 337-360.

Speier, C., Vessey, I., Valacich, J. S., 2003. The effects of interruptions, task complexity, and information presentation on computer-supported decision-making performance. *Decision Sciences, 34*, 771-797.

Waage, J. K. (1979). Foraging for patchily distributed hosts by the parasitoid Nemeritus canescens. *Journal of Animal Ecology, 48*, 353-371.

Whitman, M. E., 2003. Enemy at the gate: threats to information security. *Communications of the ACM, 46, 8*, 91–95

Wild, P.J., Johnson, P., Johnson, H., 2003. An hour in the life: Towards requirements for modelling multiple task work. *Proceedings of the ACM Conference on Human Factors in Computing Systems*, *CHI '03*, Fort Lauderdale, USA, 1016-1017.

Willingham, D. T., 2010. Have technology and multitasking rewired how students learn? *American Educator, Summer 2010*, 23-28.

		IRM		Control	
		M	*SD*	*M*	*SD*
Number of tasks completed	Easy	15.83	4.01	16.42	3.20
	Hard	12.29	2.63	12.67	1.55
Number of documents opened	Easy	17.17	4.62	17.54	4.36
	Hard	14.33	4.41	13.67	1.83
Number of switches		11.25	5.31	10.58	5.87

Table 1. Experiment 1: Number of tasks completed and documents opened from each of the two folders in the IRM and Control conditions. Also included is the number of switches between the folders.

	IRM		Control	
	M	*SD*	*M*	*SD*
Task selection time	3.22	1.00	2.12	.48
Task selection time for uninterrupted intervals	2.76	.93		
Task selection time when uninterrupted and only 1 folder checked	2.50	.78	1.97	.43
Number of intervals both folders checked	3.63	2.93	1.83	1.55

Table 2. Experiment 1: Mean task selection times (s) accounting for interruptions and interleaving between folders.

	IRM				Control			
	Switch		No Switch		Switch		No Switch	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Task Completion Time	44.74	9.52	40.61	8.62	45.24	12.51	35.34	6.07
Current task − Last task in other folder	12.84	10.47	-7.95	21.32	8.92	5.73	-6.78	15.16
Current task − Next task in other folder	4.80	10.52	-2.89	17.98	6.96	13.82	-7.36	10.38
Current task − Next task	4.80	10.52	-4.94	7.36	6.96	13.82	4.76	3.21

Table 3. Experiment 1: Comparisons of current task completion times (s) with previous and subsequent tasks. Times are grouped by condition and whether or not they preceded a switch to the other folder.

	M	*SD*
Failure to restrict access	.46	.88
Corrected failure to restrict access	.33	.70
Incorrect email when restrict access to document	1.33	1.05
Incorrect email when requesting access	.25	.53

Table 4. Experiment 1: Errors made within the IRM condition when restricting access rights or emailing document owners to gain access to documents.

		IRM		Control	
		M	*SD*	*M*	*SD*
Number of tasks completed	Easy	17	5.27	16.83	5.34
	Hard	9.96	2.77	10.17	3.24
Number of documents opened	Easy	31.88	13.09	30.92	12.36
	Hard	30.83	11.57	27.46	8.70
Number of switches		2.38	2.18	2.13	1.30

Table 5. Experiment 2: Number of tasks completed and documents opened from each of the two folders in the IRM and Control conditions. Also included is the number of switches between the folders.

	IRM		Control	
	M	*SD*	*M*	*SD*
Task selection time	2.32	.46	1.85	.23
Task selection time for uninterrupted intervals	1.99	.38		
Task selection time when uninterrupted and only 1 folder checked	1.89	.35	1.78	.23
Number of intervals both folders checked	2.67	2.18	1.75	1.03

Table 6. Experiment 2: Mean task selection times (s) accounting for interruptions and interleaving between folders.

	IRM		Control	
	M	*SD*	*M*	*SD*
Selection time for next task opened	3.32	1.40	2.06	.42
Selection time for next task completed	8.05	6.13	4.45	2.55
Documents opened before task completed	1.12	1.09	.70	.65
Documents reopened before task completed	.70	.81	.16	.33

Table 7. Experiment 2: Post-interruption task selection means in IRM condition and at matched juncture in Control condition. Times are given in seconds.

	High Interleaving				Low Interleaving			
	Alert		No Alert		Alert		No Alert	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Total task completion time	1230	207	1236	403	901	211	852	302
Proportion comprehension tasks correct	.86	.06	.91	.07	.85	.08	.88	.07
Proportion access control tasks correct	.95	.10	.89	.18	.98	.06	.86	.27

Table 8. Experiment 3: Mean total time (s) to complete all tasks and mean proportion of tasks correct.

	High Interleaving				Low Interleaving			
	Alert		No Alert		Alert		No Alert	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Middle of subtask	11.44	3.79	9.75	1.44	2.88	2.31	1.44	1.36
Between subtasks	14	3.77	14.06	3.17	11.69	2.12	12.31	2.50

Table 9. Experiment 3: Mean number of times participants interleaved between a document and the email package in the middle of a comprehension subtask or between comprehension subtasks.

Figure 1. Task length for each of the easy and hard lists of tasks. Task length is given in number of characters that have to be entered to complete the task. Each list is contained with a folder with tasks ordered from easiest at the top to hardest at the bottom.

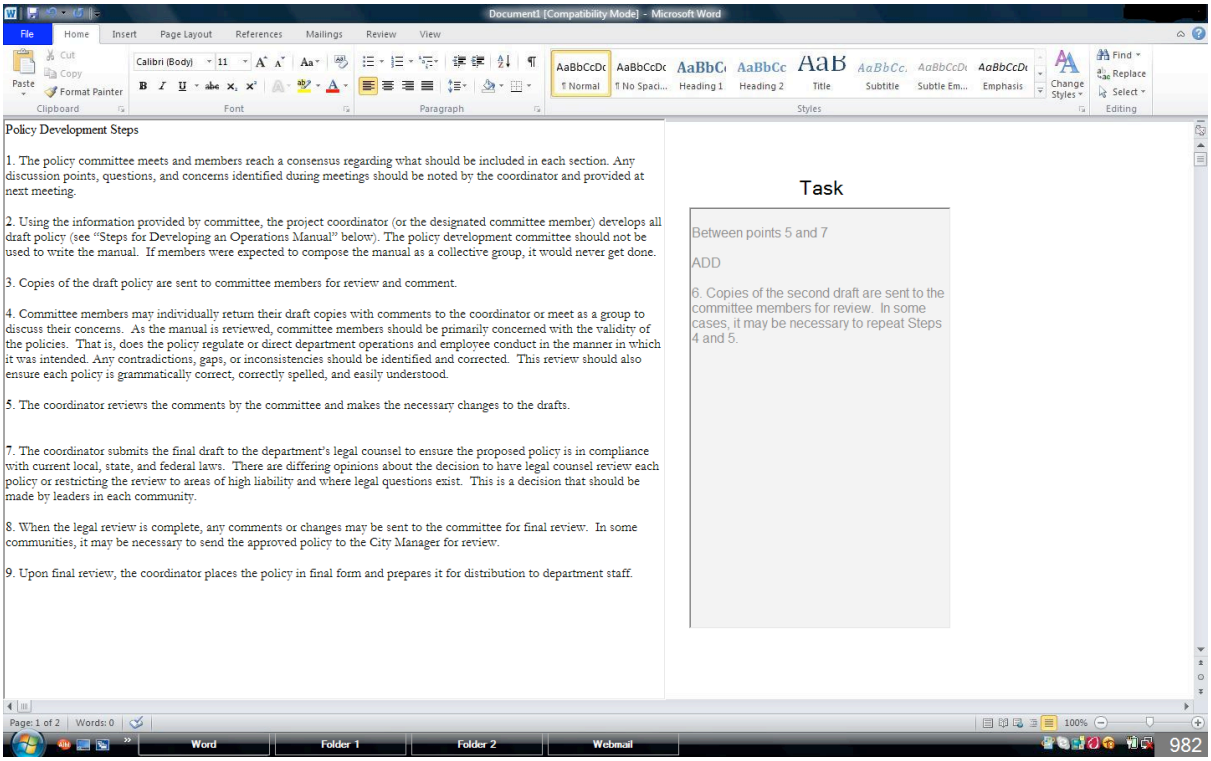

Figure 2. Screenshot from the word simulation interface.

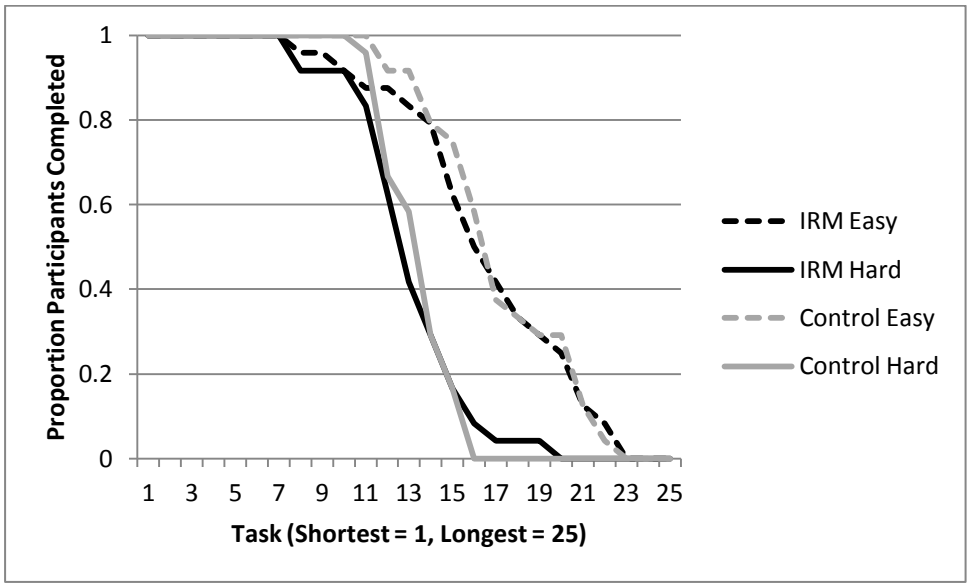

Figure 3. Experiment 1: Proportion of participants who completed each task from the easy and hard lists within each condition.

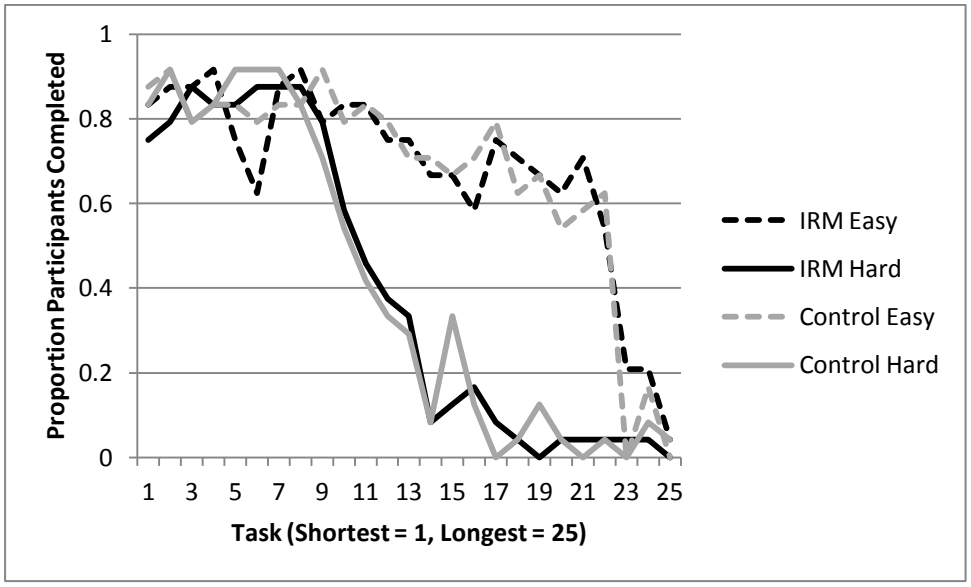

Figure 4. Experiment 2: Proportion of participants who completed each task from the easy and hard lists within each condition.