How to Mitigate Resumption Errors in a Procedural Routine after an Interruption: An Investigation into the Speed-Accuracy Trade-offs effect on Task Resumption Performance

Tianbo Xu

Project report submitted in part fulfilment of the requirements for the degree of Master of Science (Human-Computer Interaction with Ergonomics) in the Faculty of Life Sciences, University College London, 2011.

ACKNOWLEDGMENTS

Many people have contributed to the completion of this project. I would particularly like to thank my supervisor Anna Cox for her direction, input and support throughout the project period. Special thanks to Sandy Gould for his patience and generous time spent coding and revising the program for the Pharmacy-ordering task and Packing task, the experiment could not have happened without your help.

I would also like to thank all participants, especially those who found time in their busy working schedules.

My gratitude also extends to Steve Davis who provided an account to login the computer in UCL Bedford way building.

ABSTRACT

Interruptions are commonplace in most workspaces in contemporary society, which results in an increase of resumption errors, particularly, in forms of sub-task sequence errors and post-completion errors. Recovering from an interruption usually takes time, which presents as a resumption lag. This study focuses on identifying whether or not people can be motivated to reduce both sub-task sequence errors and post-completion errors via actively strategic controls, such as speed-accuracy trade-offs. The experiment in this study involved two tasks: the Pharmacy-ordering task (main task) was a computer-based procedural routine; the Packing task (interrupting task) was related to mental arithmetical problems. During the execution of Pharmacy-ordering task, participants were interrupted occasionally by the Packing task for 30 seconds. A between-subject design with three conditions, the independent variables were three levels of time-cost penalty; and main dependent variables were sub-task sequence error-rate, post-completion error-rate and resumption lag. The experiment was conducted in a cubicle located in UCL Bedford way building. Thirty-eight participants (23 females) were recruited in two ways: 20 participants were recruited from psychology subject pool and 18 were touched in UCL Bedford way building randomly. Their ages are in the range of 19 to 34. Participants were paid £3 for their time. The findings revealed people can be motivated to make speed-accuracy trade-offs in terms of sub-task sequence errors. But there was no reliable evidence to propose that post-completion errors can be motivated to reduce. The implications of this study are making people to be aware that slowing down after an interruption can trade for accuracy. Meanwhile, the findings are meaningful to improve device design.

CONTENTS

Chapter 1. Introduction 6

Chapter 2. Literature review.. 10

2.1. Resuming the primary task after an interruption 10

2.2. The effects of interruptions on sub-task sequence errors and post-completion errors 13

2.3. Speed-accuracy trade-offs theory 18

2.4. Summary.. 19

Chapter 3. Resuming the primary task with strategic controls 21

3.1. Method.. 24

Participants 24

Materials 24

Design.. 27

Procedure 29

3.2. Results 30

Overall resumption error rate.. 31

Sub-task sequence error rates and post-completion error rates 32

Task resumption lag.. 33

Relationships between error frequency and resumption lag....................... 36

Chapter 4. General Discussion 40

4.1. Relationships between sub-task sequence error rates and resumption lag 41

4.2. Relationships between post-completion error rates and resumption lag 45

4.3. Limitations.. 46

4.4. Recommendation and future experimentation 47

4.5. Implications 48

Chapter 5. Conclusion 50

CHAPTER 1. INTRODUCTION

In the contemporary world, interruptions are pervasive in most fields of working environments and have become a portion of people's daily life (Altmann & Trafton, 2004, 2007; Byrne & Bovair, 1997; Hodgetts & Jones, 2006; Li et al., 2006, 2008; Speier, Valacich & Vessey, 1999). Particularly, in some busy environments, such as public organizations and healthcare surroundings, people are frequently interrupted (Czerwinski, Horvitz & Wilhite, 2004; Gloria, Gudith & Klocke, 2008; Gonzalez & Mark, 2004; McFarlane & Latorella, 2002; Speier, Valacich & Vessey, 1999). One study, which focused on interruptions in public office workspace, showed that office workers found most of the interruptions annoying while executing a routine task. Moreover, 40% of the time, people could not recover from interruptions immediately and correctly due to the lapse of some components in the previous task or environmental context (Czerwinski, Horvitz & Wilhite, 2004). In domains with high request of reliability and safety, such as health care settings and nuclear power plants (Edwards & Gronlund, 1998), resumption errors because of interruptions may result in serious consequence. In 2000, the President of United State received a report, in which the risk of patients' safety due to incorrect operations by healthcare workers was described as a national problem (Reason, 2002). In addition, a considerable number of researches also highlighted the severity of interruptions in healthcare field (e.g., Back, Brumby & Cox, 2010; Chisholm et al., 2000; Li et al., 2006; Tucker & Spear, 2006). Consequently, identifying how people recover from an interruption to reduce resumption errors is a critical issue in the field of HCI (Mark, Gudith & Klocke, 2008; Mark & Klocke, 2008).

The issues mentioned above have caused task resumption and interruptions control to attract much attention in the field of HCI. It is well-known that interruptions have a remarkable influence on task efficiency and are likely to incur errors (Edwards & Gronlund, 1998). Some studies have achieved significant understanding of how interruptions affect a procedural routine (e.g., Altmann & Trafton, 2002; Byrne & Bovair, 1997; Li et al., 2008; Mark, Gudith & Klocke, 2008; Monk, Boehm-Davis & Trafton, 2004; Speier, Valacich & Vessey, 1999;

Trafton et al., 2003; Trafton, Altmann & Ratwani, 2009). One of the most notable theoretical frameworks is the activation-based goal memory (AGM) model by Altmann and Trafton (2002), which demonstrated the effects of interruptions and task resumption processes.

One early study represented classes of human errors (Norman, 1981). Sub-task sequence errors often occur in the process of performing a procedural task, and are caused by disorder of a task sequence. This kind of error usually includes perseveration errors, anticipation and omission errors. A perseveration error is repeating an action that has already been executed. An anticipation / omission error occurs when a user skips one or more steps during a sequence action (Altmann & Trafton, 2010; Li et al., 2008). Post-completion errors are described as a specific type of omission error or "a kind of skip" (Byrne & Bovair, 1997; Li et al., 2006). For example, a person left the original copy in the machine after completing photocopy. Li et al. (2008) demonstrated that there are equal interruption effects on post-completion errors and sub-task sequence errors. Therefore, we will focus on investigating how people deal with sub-task sequence errors and post-completion errors during the task resumption processes in one aspect of study.

In order to reduce resumption errors, some studies investigated the effectiveness of both internal and external cues on recovering from an interruption (e.g., Chung & Byrne, 2004, 2007; Hodgetts & Jones, 2006; Trafton, Altmann & Brock, 2005). The results showed that some cues are useful for people to avoid errors, such as an environmental cue (Trafton, Altmann & Brock, 2005), but some cues are not, such as an automated cue (Chung & Byrne, 2007). For a cue to be effective, it must be patently distinct. Otherwise, it may be annoying for user to operate and affect the aesthetics of user interface design (Ratwani, McCurry & Trafton, 2008). Moreover, the effectiveness of cues also depends on the characteristics of different tasks (Chung & Byrne, 2004). Some studies argued that designing appropriate user interface or redesigning the current system may help people to resume an interrupted task (e.g., Prevost et al., 2007). Particularly, for the post-completion errors, Byrne and Bovair (1997) recommend to design them out. However, designing an appropriate user interface or redesigning a device is not always possible and depends on the various features of the action context (Blandford, 2000; Ratwani, McCurry & Trafton, 2008).

It is well established that people can trade speed for accuracy (Hick, 1952; Reed, 1973; Wickelgren, 1977; Woodworth, 1899). Moreover, the theory of speed-accuracy trade-offs is widely applied in psychology and the area of human resources (Forster, Higgins & Bianco, 2003). However, there is insufficient research focusing on using this theory to interpret the relationships between how quickly people recover from an interruption and how often a resumption error occurs. If people spend relatively long time to resume an interrupted task, will they make fewer errors? By contrast, if people recover from an interruption quickly, will more resumption errors be made? What is the difference in terms of time cost between correct and incorrect task resumption? In this paper, we investigate the above questions to identify the effects of speed-accuracy trade-offs on performance of task resumption. Specifically, an experiment adapted from Li et al.'s (2008) Doughnut Machine paradigm was adopted to investigate this issue. We assert that conducting this study will be meaningful towards extending the understanding of how people conduct task resumption. Meanwhile, it is beneficial to facilitate the design of a device in the future.

This paper is organized as follows: In Chapter Two, we will review relevant literatures and summarise the previous findings, particularly, in terms of interruption disruptiveness and the task resumption processes; post-completion and other procedural errors; and speed-accuracy trade-offs theories. This will provide insights on how interruptions affect people's performance while resuming an interrupted task; and how speed-accuracy trade-offs are applied to improve task resumption. Based on these previous studies, we will consider and propose the hypotheses for this study, which we will discuss in Chapter Three. Then, we will report an experiment adopted for this study in four aspects: participants, materials, design and procedure. Lastly, we plan to illustrate the analysed data with tables and figures. In Chapter Four, based on the analysed and summarized results, we will critically discuss the main findings of this study. First, we focus on demonstrating how the findings are related to our hypotheses and previous works; and how they advance the area we are studying. Then, we propose the limitations of our study and at same time consider how they could be improved in future experiments. Moreover, we will extend current study to propose future research. Next, we discuss the implications of our findings. In the final chapter, we conclude and restate the key ideas and findings mentioned in the paper.

CHAPTER 2. LITERATURE REVIEW

In this chapter, the review of literatures offers context into investigation of how people recover from an interruption. Specifically, based on the previous studies, we will identify the effects of interruptions on a sequential action performance. In this part, we will focus on Altmann and Trafton's activation-based goal memory (2002) model to understand how interruptions affect people while resuming a task, and what characteristics of interruptions make them disruptive. Then, we will examine the similarities and differences between post-completion errors and sub-task sequence errors, which may help us to decide whether we may investigate sub-task sequence errors and post-completion errors in the same study; finally, we will investigate how the theory of speed-accuracy trade-offs assists people to conduct task resumption.

2.1. Resuming the primary task after an interruption

In order to understand how people recover from an interruption, first and foremost, it is necessary to identify the process of task resumption. Altmann and Trafton's activation-based goal memory (AGM) model (2002) explained the goal encoding and memory retrieval in great detail and depth. Moreover, it has been applied to predict the crucial elements which dominate the task resumption processes (Trafton et al., 2003).

Among the earlier studies, ACR-T cognitive theory (Anderson and Lebiere, 1998) analyses the activations during the process of reaching a goal. Based on the ACT-R theory (Anderson and Lebiere, 1998), Altmann and Trafton (2002) proposed the AGM model to explain the activations of memory items and provided a method to analyse goal encoding and retrieval. According to the AGM (Altmann & Trafton, 2002) model, interruptions result in memory decay for suspended goals. Altmann and Trafton proposed that people can respond to interruptions slowly, so that they can use the duration of time between the occurrence of an interruption and the start of response called "interruption lag" to rehearse the related information of the primary task (Altmann & Trafton, 2002). Furthermore, another study indicated that longer interruption lag can better assist people to recover from an interruption (Trafton et al., 2003). However, interruptions often occur immediately and urgently, which do not allow people to control the response with strategies. Under this situation, an interruption prevents or minimizes the information rehearsal, so it is prone to make resumption errors (Altmann & Trafton, 2002). A possible solution, as predicted by the AGM model, is that people can rehearse suspended goals during an interruption.

In the AGM (Altmann & Trafton, 2002) model, they argued that memory always returned the most active item at that moment when people resumed the primary task. But Altmann and Trafton also proposed that the AGM model was flexible for people to retrieve other memory items, which depends on time course and memory decay. If a particular item is required to capture, the memory system will analyse the activation of the previous item using trace and current mental or environmental context (Altmann & Trafton, 2002). In contrast to the principle of stack that is based on a “last-in, first-out” rule, the AGM (Altmann & Trafton, 2002) model proposes that there is no fixed order. Memory retrieval for goals is dependent on activation level at that instant. Previous goals are stored at an “interference threshold” (Altmann & Trafton, 2002), which is described as “mental clutter”. The suspended goals can be retrieved only when their activation level is beyond the threshold. The amount of time people need to engage in resuming the suspended goals depends on the level of activation above threshold. That is why people spend relatively longer time to resume some suspended goals than others. The results of the “Tower of Hanoi” experiment conducted by Altmann and Trafton (2002) showed that it usually took approximately at least one or two seconds to raise activation and then retrieve the memory. Moreover, in the experiment, they simulated resumption lag and errors by applying the AGM model, and found that problem solving is closely related to interruptions and task resumption.

In the AGM model, Altmann and Trafton (2002) predicted that the activation level will decrease with increasing length of an interruption. Anderson and Douglass’s model (2001) also support that the duration of an interruption is related to task resumption. A study by Peterson and Peterson (1959) proposed that the memory for goals will decay over time. Moreover, it was more difficult to retrieve and the likelihood of making an error increased with decay of memory. However, conversely, in Gillie and Broadbent's study (1989), they used a computer-based shopping task with two durations of interruptions that lasted 30 seconds and 2.75 minutes, respectively. This was done to clarify the effects of interruption length. The results revealed that the duration of an interruption alone was not an important factor that determined whether an interruption was disruptive or not. Similarly, another study by Bailey, Konstan and Carlis (2000) focused on investigating interruptions during routine activities, such as reading. The study also proposed that the length did not make an interruption significantly disruptive. Until the recent studies, Hodgetts and Jones (2006) provided a further understanding about the effects of interruption length. Based on the classic Tower of Hanoi task, they developed a task called Tower of London, in which participants were required to move five different-coloured disks among three pegs to match the configuration shown on the screen. During the movement, participants were interrupted by a mood checklist task with two types of duration: six seconds and 18 seconds. The results verified Altmann and Trafton's assumption (2002) and indicated that with the increasing interruption length, people experienced a longer resumption lag for interruption recovery. However, the Tower of London task did not yielded significant difference in terms of error rates between the interruption condition and control condition. Hodgetts and Jones predicted that this might suffer from the limitations of interrupted task complexity (Hodgetts & Jones, 2006). A study conducted by Monk, Trafton and Boehm-Davis (2008) identified the main reasons why some studies (e.g., Cellier & Eyerolle, 1992; Czerwinski et al., 2000; Gillie & Broadbent, 1989; McFarlane, 2002; Oulasvirta & Saariluoma, 2004; Zijlstra et al., 1999) failed to identify the effects of interruption length on resumption lag. One reason was that the studies did not adopt sufficiently perceptive ways to test how fast people conduct task resumption followed by an interruption. The other cause was inappropriate interruption length was involved in the studies. For instance, the interruption duration in Gillie and Broadbent's study (1989) was 30 seconds and

2.75 minutes, respectively. In light of the AGM (Altmann & Trafton, 2002) model, the memory decays over time. However, after a certain period of time, the decay is presented as an asymptotic level. Therefore, if 30 seconds have arrived at that asymptotic area, interruptions for 30 seconds and 2.75 minutes had no significantly different effects on activation level. Meanwhile, they enhanced the findings derived from Hodgetts and Jones (2006). In the study, participants were asked to complete a simulated VCR task including four sub-tasks ("show's starttime", "end-time", "day of week", and "channel number") with clicking according buttons. In the midway, theVCR task was interrupted by a pursuit tracking task for three types of duration (3s, 8s and 13s). The results showed that the resumption lag of overall interruption conditions was longer than control trials, which indicated the basic disruptive effects of interruptions on resumption lag. Furthermore, the resumption lag increased with the increasing length of interruption duration, which was consistent with Hodgetts and Jones's findings (2006). Meanwhile, the results also showed that the errors rates of the three interrupted conditions did not make remarkable difference. In other words, the study did not find reliable evidence that reveals the effect of interruption length on resumption error rates (Monk, Trafton and Boehm-Davis, 2008).

2.2. The effects of interruptions on sub-task sequence errors and post-completion errors

As mentioned above, some studies have provided sufficient understanding regarding the effects of interruption length on resumption lag. However, there are no significant findings that are related to the effects of interruptions on resumption error rates both in the problem solving tasks and well-learned routine procedures (Hodgetts & Jones, 2006; Mark, Trafton & Boehm-Davis, 2008). To some extent, this situation is due to the very low resumption errors rates which occurred in the studies. This increases the difficulty to identify the exact effects of an interruption on resumption errors. In order to improve this situation, some studies developed paradigms in order to generate relatively high resumption error rates for the laboratorial studies, which assisted with further research on interruptions (e.g., Back, Brumby & Cox, 2010; Byrne & Bovair, 1997; Li et al., 2006, 2008; Trafton, Altmann & Ratwani, 2009).

Sequence errors are common procedural errors that occur during a routine sequential task. In light of the classification of errors proposed by some previous studies (e.g., Norman, 1990; Reason, 1984), sequence errors are composed of perseveration errors, anticipation / omission errors and intrusion errors. Specifically, a perseveration error refers to repeating a previous step. Moreover, perseveration errors usually include two sub-types, continuous perseveration errors and recurrent perseveration errors. Conversely, anticipation and omission errors mean skipping one or more actions during a sequence (Trafton, Altmann & Ratwani, 2009). Based on the AGM (Altmann & Trafton, 2002) model, Trafton, Altmann & Ratwani (2009) developed the memory for goals (MFG) model to predict the effects of interruptions on sequence error rates during a procedural task. The MFG model demonstrated why perseveration errors and anticipation / omission errors are prone to occur in a noisy system. Specifically, the interruptions as a function of noise in a procedure may lead to the recurrence of a completed episodic memory, and then perseveration errors occur. Anticipation / omission errors are due to the disorder of completion monitoring, thus one or more actions are skipped. Moreover, the model also assumed that perseveration errors were more prone to be triggered than anticipation / omission errors (Trafton, Altmann & Ratwani, 2009).

Trafton, Altmann and Ratwani (2009) conducted a couple of experiments with a sequential paradigm that is adapted from Li et al.'s Doughnut Machine task (2008). This is because Li et al. (2008) proposed that the possibility of making an error is higher even for skilled users and well-learned tasks due to interruptions. Moreover, the task without global position tracing ensures that the next action of the task is not available to be identified by cues. The first experiment in Trafton, Altmann and Ratwani's study (2009) required participants to complete a task called "Sea vessel" which included five sub-tasks in a fixed order. During the execution, interruptions may trigger the moment participants complete one of the sub-tasks, and participants were asked to answer a set of arithmetical questions. The results showed that the error rate of the control condition (without interruptions) was merely 0.9%, while the error rate of the interruption condition was 9.3%. The extremely low error rate in the control condition revealed that the "Sea Vessel" task was a well-learned task, and a much higher error rate in interruption condition confirmed the effects of interruptions on sequence error rates. The results are consistent with Li et al.'s demonstration (2008). Moreover, the findings also showed that perseveration errors occurred more commonly than other types of sequence errors.

Compared with the first experiment by Trafton, Altmann and Ratwani (2009), the second experiment was more complicated, in which participants needed to perform a computer-based sequence task in relation to stock deals. Moreover, detailed information was required of the task, such as prices, changed over time. The interrupting task was similar to the one in Experiment 1. The results enhanced the findings obtained in the first experiment. The error rate in the interruption condition was 6.7% comparing to 0.3% in the control trials. The perseveration errors accounted for 60.6% of all sequence errors.

Trafton, Altmann and Ratwani's research (2009) explicitly demonstrated the effects of interruptions on various types of sequence errors. In a well-learned procedural action without visible cues, the error rates were significantly higher due to interruptions. The AGM model (Altmann & Trafton, 2002) and MFG (Trafton, Altmann & Ratwani, 2009) model clarified the effects of an interruption in terms of time costs and sequence error rates, respectively. However, both of these studies (Altmann & Trafton, 2002; Trafton, Altmann & Ratwani, 2009) lack demonstration in the relationship between time costs and error rates, particularly, the analysis and comparison of resumption lag for correct and incorrect task resumption.

Post-completion errors are described as a type of omission error. When a main goal is completed, an extra action named post-completion step may be skipped, causing post-completion errors to occur. A landmark study related to post-completion errors was conducted by Byrne and Bovair (1997). They elaborated that working memory was an important factor that induced post-completion errors. In the study, Byrne and Bovair adopted a between-subjects design with three factors: "working memory load" (no load and load), "working memory capacity" (low capacity and high capacity) and "mix of tasks" (post-completion / control version of "Phaser" task and post-completion / control version of "Transporter" task). The concurrent task was unrelated to the sequential tasks, in which participants were required to remember the letters they heard during the period of 9-45 seconds. The concurrent task was designed for measuring the working memory load. The results showed that in the condition of high working memory capacity without working memory load, the post-completion errors were few. In the condition of low working memory capacity with working memory load, the number of post-completion errors significantly increased. However, there is no reliable distinction between high working memory capacity with working memory load condition and low working memory capacity without working memory load condition (Byrne & Bovair, 1997).

In order to obtain insights into post-completion errors, Li et al. conduct a further investigation to indentify the effects of interruption positions and length on post-completion errors (2006). In the first experiment of Li et al.'s study (2006), participants were asked to perform a computer-based Doughnut Order task which was made up of two parts: the doughnut-making task and order collection task. The doughnut-making task has five components in a specified order and a post-completion step which is labelled "Process / Cleaning". The order collection task named "Call Centre" simulated the real workspace where it is reasonable to collect an order away from the doughnut making task. During the execution, the interrupting task that lasts for 75 seconds might be activated immediately after one of the sub-tasked is completed or just before the post-completion step. During the interrupting task, participants were required to calculate a series of mathematical questions. According to the results, when the interruptions occurred just before the post-completion step, the post-completion error rate is 29.8%. The post-completion error rates in control condition and the condition that interruptions were triggered immediately after one of the sub-task completion were 8.6% and 8.9%, respectively. Subsequently, the statistic analysis verified that the interruption positions had significant effects on post-completion error rate.

Compared with experiment 1, experiment 2 in Li et al.'s study (2006) used the same task, but classified the interruption duration (15 and 45 seconds). The findings of experiment 2 in terms of interruption positions strengthened the results in experiment 1. But the effects of interruption length on post-completion error rates were mild. Consistent with the studies mentioned above (Hodgetts & Jones, 2006; Monk, Trafton & Boehm-Davis, 2008), there is no strong evidence to reveal the effects of interruptions on error rates.

In order to identify the relationship between post-completion errors and other sequential errors, a study was conducted by Li et al. (2008). They adopted the same paradigm as the one used in Li et al.'s study in 2006. In experiment 1, they tested the interruption position effects on both post-completion errors and other procedural sequence errors (sub-task sequence errors and sub-task initialization errors). The results of post-completion errors confirmed the findings of their previous study in 2006. The results of sub-task sequence errors showed that these errors were made frequently (22.6%) when the interruptions occurred immediately after one of the sub-tasks. The sub-task sequence error rates were significantly low later after the interruptions (1.1%). In the situation without interruptions, the sub-task sequence error rate was 0.6%. Moving to the results for sub-task initialization errors, it was found that there was no significant interruption position effect on these types of sequence errors (Li et al., 2008). To summarise the findings, we can see sub-task sequence errors are significantly affected by interruption positions.

Experiment 2 replicated experiment 1, but the interruptions occurred less frequently, which aimed to investigate the interaction between post-completion errors and sub-task sequence errors. The results enhanced the findings in experiment 1. Furthermore, it was found that more post-completion errors were made when interruptions were not instantly triggered (Li et al., 2008).

According to the studies reviewed above, we identified the interruption position effects on both post-completion errors and other procedural errors. Furthermore, it is verified that there is no significantly different effect of interruptions on post-completion errors and sub-task sequence errors (Li et al., 2006, 2008). This is encouraging evidence towards investigating both post-completion errors and sub-task sequence errors in one aspect of the study. Moreover, the findings of interruption position effects propose an appropriate way to design our experiments.

2.3. Speed-accuracy trade-offs theory

Interruptions do have effects on post-completion and sub-task sequence error rates (Li et al., 2006, 2008; Trafton, Altmann & Ratwani, 2009). Therefore, it is necessary to identify effective ways to mitigate these errors. A relative study was conducted by Chung and Byrne (2004), in which the visual cue and the mode indicator were involved in the experiment to test whether or not they were effective in reducing post-completion errors. The results showed that the visual cue could reduce post-completion errors effectively. However, integrating visual cues into a system leaves some unconfirmed factors. For example, how the individual strengths of visual properties in terms of cue timing, cue movement and cue meaningfulness are decided (Chung & Byrne, 2004). Moreover, this study only focused on the cue effectiveness on post-completion errors, which is not clear how to mitigate other procedural errors in a routine action.

It has been well established that people can trade speed for accuracy in various activities. Some early studies focused on the speed-accuracy trade-offs theories in terms of recognition memory retrieval (e.g., Boldini, Russo & Avons, 2004; Reed, 1973). In the studies, they proposed that the response time delay for a short period of time was significantly effective in improving accuracy levels for retrieving recognition memory, but the accuracy would not further increase when the response delay was extended for a long period time. Therefore, the researchers predicted that there may be an optimal response delay to obtain best accuracy.

A recent study by Back, Brumby and Cox (2010) focuses on the relationships between resumption lag and error rates. In the study, they adapted the paradigm by Li et al. (2008): an enforced delay condition was introduced. In the condition, after interruptions, the interaction between participants and the main task (doughnut-making) was locked out for 10 seconds. During this lockout time, participants were able to view the interface, but could not operate it. The results indicated that an enforced locked-out condition reduced sequence errors dramatically (only 36% of the errors made in the interruption-only condition). Furthermore, by analysing resumption lag in interruption-only condition, significantly more sequence errors were made when resumption lag was shorter than 2.5 seconds. By contrast, few errors were made when the resumption lag was longer than 2.5 seconds. Although this study did not explain the data using the MFG (Trafton, Altmann & Ratwani, 2009) model, it did reveal that speed-accuracy trade-offs can be used strategically to reduce sequence errors while recovering from an interruption (Back, Brumby & Cox, 2010). However, this study only investigated the relationship between resumption speed and sequence errors rates. Therefore, the relationship between resumption lag and post-completion error rates needs further investigation.

2.4. Summary

In brief, based on the activation-based goal memory (Altmann & Trafton, 2002) model, we obtained insights into the task resumption processes. Moreover, we identified what characteristics of interruptions contribute to their disruptiveness via a set of classic studies, such as the interruption duration effect on resumption lag (e.g., Gillie & Broadbent, 1989; Hodgetts & Jones, 2006; Monk et al., 2008). Subsequently, we reviewed a number of studies (e.g., Byrne & Bovair, 1997; Li et al., 2006, 2008; Trafton, Altmann & Ratwani, 2009) to clarify the effects of an interruption on both post-completion errors and other procedural errors. For instance, the studies demonstrated the effects of interruption positions and length on post-completion errors and sub-task sequence errors. Moreover, the extensive experiments conducted in the studies informed the current study on which paradigm would be suitable to effectively conduct our study. Finally, we investigated the relevant studies in terms of speed-accuracy trade-offs theories (e.g., Back. Brumby & Cox, 2010; Chung & Byrne, 2004; Reed, 1973), which made us aware of the current findings. By critically reviewing the studies, we extended and deepened our understanding in terms of how to recover from an interruption. Meanwhile, we need to strengthen insufficient aspects of the studies.

CHAPTER 3. RESUMING THE PRIMARY TASK WITH STRATEGIC CONTROLS

The experiment described in this chapter is based on the classic memory for goals model in terms of interruption study (Altmann & Trafton, 2002; Trafton, Altmann & Ratwani, 2009) and adapted from a widely applied doughnut-machine paradigm (Li et al., 2006, 2008), whereby participants executed a procedural routine with occasional interruptions. It is designed to explore whether or not the speed-accuracy trade-offs theories can be appropriately applied as a strategy while recovering from an interruption. Specifically, we would like to identify the relationship between how quickly people recover from an interruption and how frequently post-completion errors and sub-task sequence errors may be made. In other words, two factors attracted our interests: resumption lag and frequency of both post-completion errors and sub-task sequence errors.

Two tasks were introduced in the experiment: the Pharmacy-ordering task (the main task) and the Packing task (the interrupting task). The main task was a sequential routine, in which participants were required to complete a prescription sheet in specified order. The interrupting task involved mental arithmetic. A participant might be interrupted by the Packing task while executing the Pharmacy-ordering task. The independent variable was system time-cost penalty when participants worked on an incorrect sequence after an interruption. In the experiment, three levels of penalty period were set: low level, medium level and high level. In the low level condition, the participants were not penalized when they resumed the primary task incorrectly. In the medium and high level, the task interface was blocked by an error page for 20 and 40 seconds respectively when a resumption error was made. The main dependent variables were resumption lag and both frequency of sub-task sequence errors and post-completion errors. The entire experiment lasted approximately 45 minutes. The brief process was signing a consent form, listening to the introduction of the tasks, exercising and executing the trials. Finally, a verbal debrief was optional for participants. Participants were paid £3 for their time.

As mentioned above, in the low condition, there was no penalty, meanwhile, a reminder message could aid participants to work on the correct action when they made resumption errors. In the medium and high level conditions, time-cost penalty triggers the moment participants inaccurately resume the primary. Based on common sense, it is reasonable to predict that participants who executed the primary task in the low condition might not consider the consequence of making resumption errors. In other words, they might work on task resumption by intuition or quick guess. For the participants in the medium and high level condition groups, the time-consuming penalty drove them to conduct the task resumption processes more discreetly. Moreover, the penalty duration in the high level condition was longer than in the medium condition. Participants were motivated to obtain a more confident answer before task resumption in the high level condition. In other words, further costly penalty strengthened the motivation of participants to avoid resumption errors. Therefore, across the three conditions, we predict that participants would take the longest time to recover from an interruption in the high condition, and participants in the low level condition would resume the primary task most quickly.

In order to avoid the penalty, the people in the medium and high condition groups might adjust retrieval strategy. On one hand, actively appropriate retrieval strategies might improve the performance of task resumption. For example, a possible solution proposed by the AGM (Altmann & Trafton, 2002) model was that people could rehearse the suspended goals before or during the interruptions. Rehearsals could assist memory retrieval and strengthen activation of suspended goals. To some extent, the activation level determined whether or not a suspended goal could be retrieved. On the other hand, the speed-accuracy trade-offs studies in terms of cognition memory revealed that longer response times can improve accuracy (Boldini, Russo & Avons, 2004; Reed, 1973). Moreover, in light of Li et al.'s studies (2006, 2008), it was confirmed that there is no significantly different effects of interruptions on post-completion errors and sub-task sequence errors when interruptions occurred on specific positions. Therefore, we predict that both sub-task sequence error rates and post-completion error rates in medium and high level condition would be lower than the rates in the low level condition.

However, there is ambiguity in the prediction of whether or not resumption errors would be further reduced in the high level condition. An aspect of speed-accuracy trade-offs study indicated that continuous increase in response time did not significantly contribute to accuracy. Moreover, the study also predicted that there might be an optimal response time for best accuracy (Reed, 1973). In light of the AGM (Altmann & Trafton, 2002) model, the suspended goals could be retrieved only at the moment their activation level was beyond the threshold. The length of time people took to resume the suspended goals depended on how far the activation level was above threshold (Altmann & Trafton, 2002). It is difficult to judge the causes for resumption errors in the medium level condition, in which the errors might be due to insufficient resumption lag (i.e. the duration of resumption lag has not reached the optimal length.) or low activation level. In other words, if most of the errors made in the medium level condition were due to insufficient resumption lag, longer resumption lag in the high level condition could reduce the errors further; if most of the errors were due to low activation level, the longer resumption lag was no use. Therefore, based on the above reasoning regarding, about the error rates for sub-task sequence and post-completion errors in the medium and high level conditions, we propose a pair of conflicting predictions.

In brief, we summarize the hypotheses based on the relationship between resumption lag and error frequency, as follows:

- Participants in the low level condition tended to recover from an interruption quickly. In the medium level and high level conditions, people were motivated to take relatively longer resumption lag to retrieve rehearsal processed and make decision in order to reduce both post-completion errors and sub-task sequence errors. Moreover, resumption lag in the high level condition would be further longer than the resumption lag in medium level condition.

- Both sub-task sequence and post-completion error rates in medium and high level conditions would be lower than the rates in low level condition. But about the error rates between medium and high level conditions, there are two conflicting predictions: 1) the error rates in the high level condition are lower than the rates in the medium condition; 2) the error rates in both conditions are not significantly different.

3.1. Method

Participants

Twenty participants were recruited from the psychology subject pool, and 18 participants were recruited in the lobby of UCL Bedford Way building. The 38 participants (23 were female) came from varied academic backgrounds, excluding psychology, and had a mean age of 25.6 years (SD = 3.3). All were allocated one of the conditions randomly (low level: 12; medium level: 12; high level: 14). The payment for participants was £3.

Materials

The two tasks used in this study were adapted from the classic Doughnut machine paradigm (Li et al., 2008). One was the primary task, called Pharmacy-ordering task; the other was used as an interrupting task, named Packing task.

The primary task was a sequential action called the pharmacy-ordering task. The goal was to enter three groups of information related to a pharmacy order in a prescription order in a pharmacy (Figure 1). Prescription details were displayed in the centre of the "Prescription Machine" screen the moment "Next prescription" button was clicked, and the correct sequence was specified as "Type", "Shape" and "Colour". In order to complete each sub-task in a trial, participants have to first click the according button in the selector area that is situated on the right side of the task screen, and then enter the related numbers that were shown in the first column of the prescription sheet. The final step to complete the sub task was clicking the "OK" button at the bottom of each section's area. When participants moved to the following sub-task, the entered numbers would be eliminated. This is because we would not provide any visual cues to assist participants to decide what to do next. Subsequently, participants need to click the “Process” button that was located at the lower right hand corner of the screen to submit the order. After clicking the “Process” button, the button label changed to “Achieve” automatically. Then, the last action for the entire process was clicking the “Achieve” button to save the order. To summarise the whole procedure of each trial, it is along with a linear sequence: Next Prescription, Type, Shape, Colour, Process and Achieve.

Figure1. Screenshot of the Pharmacy-ordering task

The Packing task was an interrupting task, in which participants were required to answer a series of mental arithmetical questions. The information relevant to the question was displayed in the centre of the task screen including the amount and type of prescription and blister pack capacities (Figure 2). In this task, the amount and type of prescription were distributed randomly, and the blister pack capacities were four and nine constantly. The question rule was calculating how many four and nine capacity blister packs to right accommodate the prescription in the required amount (i.e., 4 x 4 capacity packs plus 1 x 9 capacity pack was 25 exactly). Participants are asked to input appropriate numbers in the two textboxes respectively to match the question rule, and then click the “Pack” button to submit the answer. Once participants completed a question, the prescription amount and type automatically changed to another.

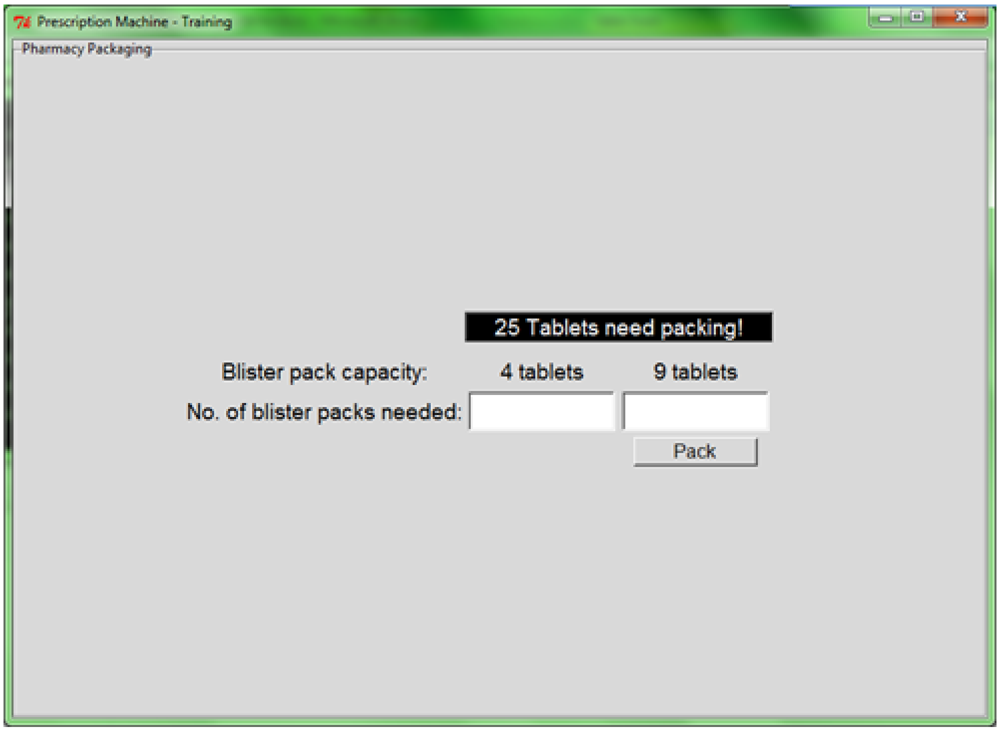


Figure2. Screenshot of the Packing task

During executing the Pharmacy-ordering task, participants might be interrupted by the Packing task occasionally for a 30-second period immediately after they completed one of the sub-tasks or just before clicking on the “Achieve” button. To be more specific, the exact points that interruption might occur in the entire trial sequence were illustrated in Table 1.

Table 1
The entire task sequence and interruption points

Step number	Step name	According operations
1	Next Order	Click on the "Next Prescription" button to show order details
2	Select_Type	Click on the "Type" button on selector
3	num_Type (x)	Enter numbers in the appropriate textboxes of Type compartment
4	num_Type (y)	Enter numbers in the appropriate textboxes of Type compartment
5	OK_Type	Click on OK button in the Type compartment
Interruption P		
6	Select_Shape	Click on the "Shape" button on selector
7	num_Shape (x)	Enter numbers in the appropriate textboxes of Shape compartment
8	num_Shape (y)	Enter numbers in the appropriate textboxes of Shape compartment
9	OK_Shape	Click on OK button in the Shape compartment
Interruption Q		
10	Select_Colour	Click on the "Colour" button on selector
11	num_Colour (x)	Enter numbers in the appropriate textboxes of Colour compartment
12	num_Colour (y)	Enter numbers in the appropriate textboxes of Colour compartment
13	OK_Colour	Click on OK button in the Colour compartment
Interruption R		
14	Process	Click on the Process button to submit the order
Interruption Z		
15	Achieve	Click on the Achieve button to save the order (Post-completion step)

The paradigm including both Pharmacy-ordering task and Packing task was programmed with Visual Basic 6 and ran on an ACER Aspire 3820T laptop. The typical performance of the laptop was Inter Core i5 CPU, 4GB memory, 500GB hard disk and 13.3 inches screen set at a resolution of 1366 x 768. The operation system is Windows 7 Home Premium 64-bit.

The whole experiment was held in the cubicle of UCL Bedford way building. The room was an enclosed space without windows, measuring approximately three square metres. There was a desk and a swivel chair in the cubicle. A set of documents were prepared in advance to assist the experiment including a consent form, a receipt form and an information sheet for the task operations. Coins were changed in order to make payments to the participants.

Design

A between-subject design with three conditions (the low level condition, the medium level condition and the high level condition) was adopted in this experiment in order to identify the differences of both resumption lag and error frequency under various resuming strategies. In the low level condition, when participant resumed the primary task on an incorrect compartment after interruptions, an error message in the lower left of task screen would remind participants what they should have done, but there is no further penalty. Participants needed to correct operations in light of the information shown in the error message. In the medium level condition, when participants recovered from interruption on an incorrect sequence, a 20-second period time-cost penalty would by triggered, during which, the interaction of task interface was unavailable. Moreover, the task interface screen changed to an error page. After the penalty period, participants had to restart the trial once again. Moreover, there was no interruption on the repeated trial to ensure that all participants receive the same number of interruptions during the entire experiment. When a resumption error was made in the high level condition, participants would suffer from the similar situation as in the medium level condition. However, the penalty was more costly. Specifically, the duration was 40-seconds. Among the three conditions, anytime when the participants made an error before the next interruptions, the system showed a message to reveal the correct action. The system recorded all operation traces for each participant in a data file.

The independent variable was time-cost penalty when participants worked on an incorrect sub-task after an interruption. The control condition was the low level condition without penalty on resumption errors. The main dependent variables were how long participants took to recover from an interruption and how frequently they made post-completion errors and sub-task sequence errors, respectively. “Achieve” was defined as a post-completion step in this experiment. A post-completion error was working on an inappropriate action when participants recovered from the interruptions that occurred just before the post-completion step. The resumption errors due to the interruptions that occur immediately after one of the sub tasks were sub-task sequence errors.

The experiment was composed of three parts. At first, participants were provided with an opportunity to practice both the primary and interrupting task with five trials. Among the five trials, three were non-interruption; one was an interrupted trial with an interruption after one of the sub-tasks; one was interrupted trial with an interruption before the post-completion step. Then, the experiment started formally. There were 27 formal trials in total that needed to be executed with an equal distribution of interruption occurrence (non-interruption: nine trials; interruptions after one of the sub-tasks: nine trials; interruptions before post-completion step: nine trials). In the midway of formal testing part, the system transferred to a named “Break” page. It is optional for participants to have a five-minute break, and then the system returned to task interface automatically. Alternatively, participant could click on the “Skip” button that is shown in the upper centre of the page to resume the test anytime within the five-minute period.

Procedure

At the beginning, participants are aware that they needed to execute a procedural routine with occasional interruptions and we expected to identify how people deal with the interruptions and restart the interrupted task. After we introduced the purpose and motivation of this experiment, participants were asked to read the consent form carefully which is approved by UCL Ethics Committee with project ID number. If they had no dissent, they would sign their names and current date on the form.

Subsequently, they were assigned one of the three conditions randomly. Based on the according condition, a verbal task manual explanation associated with a printed out task manual diagram were given to participants. The difference among the description for the three conditions was the varied consequence after a resumption error. Next, in order to make them to familiarize the operations further, participants were offered a training session, in which participants were required to execute five trials to improve the understanding on both the Pharmacy-ordering task and the Packing task.

Formal test started after training. Participants are informed that there were 27 trials in total across the study and when they executed the primary task, interruptions that lasted 30 seconds might occur occasionally immediately after the completion of one of the tasks or just before post-completion step. There were nine trials including interruptions that occurred immediately after one of the sub-tasks; nine trials including interruptions that occurred just before post-completion step and nine trials without interruptions. However, the interruption order was random. During the interruptions period, participants needed to conduct the Packing task and were encouraged to calculate as many arithmetical problems as they could. In the halfway of the forma testing, participants could choose to have five-minute break flexibly.

After executing the final trials, participants signed a piece of receipt form and were paid £3 for their time. The entire process of the experiment lasted approximately 45 minutes.

3.2. Results

The experiment aimed to identify how people perform task resumption with active rehearsal and memory retrieval strategies under the situations that varied levels of penalties triggered after a resumption error. The resumption errors attracted our attention in this study included sub-task sequence errors and post-completion errors. A sub-task sequence error was resuming the Pharmacy-ordering task at incorrect sequential points after an interruption that occurred immediately after one of the subtasks (Type, Shape and Colour). A post-completion error was affirmed when people recovered from an interruption that occurred just before post-completion step (Achieve) on inappropriate action. Three independent variables were involved into this study: sub-task sequence error rate, post-completion error rate and task resumption lag.

There were 38 participants in total recruited to collect data in the experiment. In both low level and medium condition, there were 12 participants; 14 participants perform the high level condition. However, the data from three participants was moved out from further analysis. Two of them who were assigned the low level condition and high level condition respectively made both sub-task sequence errors and post-completion errors almost every trial. Moreover, they occasionally waited for over one minute to restart the interrupted task. Therefore, we estimated that they might not understand the task sufficiently. The other participant quitted the experiment in the midway due to impatience.

Overall resumption error rate

Across the 35 participants, merely two participants did not make any resumption errors. To be specific on each type of errors: only five people did not make sub-task sequence errors and over half of the participants (22 out of 35) made at least one post-completion error. Table 2 shows the general distribution of resumption errors and overall resumption error rates across the three conditions. A total of 110 resumption errors were made across 630 opportunities (35 participants * 18 one-interruption trials). Nearly half of all the resumption errors were made in the low level condition (52 out of 110). In medium and high level condition, the numbers of resumption errors were 35 and 23, respectively. The mean error rates across the three conditions were 26.77% (53 out of 198), 16.20% (35 out of 216) and 10.65% (23 out of 216). Seventy-nine subtask sequence errors and 33 post-completion errors consisted of all the resumption errors. Specifically, there were 37 sub-task sequence errors and 15 post-completion errors made in the low level condition. The number of both types of errors presented a declined trend in the medium level condition, 25 sub-task sequence errors and 10 post-completion errors were made respectively. In the high level condition, the number of sub-task sequence errors and post-completion errors were further reduced to 17 and six, respectively.

Table 2
Overall resumption error rates across the three conditions

	low level	medium level	high level	Total no.
sub-task sequence errors	37	25	17	79
Post-completion errors	15	10	6	31
Total no. of resumption errors each condition	52	35	23	110
Total no. of opportunities each conditions	198	216	216	630
(no. of paricipants x interruption trials)	(11 x 18)	(12 x 18)	(12 x 18)	
Overall mean error rate	26.26%	16.20%	10.65%	

Sub-task sequence error rates and post-completion error rates

Figure 3 illustrates both sub-task sequence error and post-completion error rates across the low level, medium level and high-level conditions.

Firstly, in terms of sub-task sequence errors, the error rate was extremely high in the low level condition, in which 37 errors were detected across a total of 99 opportunities (M=37.3%, SD=22.3%). In medium level condition, the number of errors reduced by approximately one third, there were 25 errors made across the 108 opportunities (M=23.1%, SD=18.0%). It was observed that the error rates in the high level condition mitigated slightly than the error-rate in medium level condition (17 out of 108 opportunities, M=15.7%, SD=11.0%). Across the three conditions, although the overall error rates presented as a declined trend, we cannot judge whether or not there was reliably different error-rates across the three conditions. In order to identify the significant effect of error rates across the three conditions, a one-way ANOVA was conducted for sub-task sequence errors, $F(2, 32) = 4.44$, $p<0.05$, which showed that there was reliable difference between at least two conditions across the three conditions. Furthermore, Tukey HSD post-hoc comparison was planned. The results showed the error-rate in the low level condition was significantly higher than the error rates both in the medium and high level conditions (p's<0.05); but no reliable difference in terms of the medium level versus high level ($p>0.05$).

On the other hand, the frequency of post-completion errors across the three conditions indicated the similar tendency with the situation of sub-task sequence errors. There were 15 post-completion errors made in low level condition, and the error-rate was highest across the three conditions (M=15.2%, SD=9.0%). There were 10 post-completion errors in medium level condition and the mean error rate decreased to 9.3% (SD=9.3%). Compared with the medium level condition, the error rate reduced further (M=5.5%, SD=7.5%) and there were only six post-completion errors made out of 108 opportunities. However, according to the ANOVA ($F(2, 32) = 2.74$, $p=0.08$), we can see although there was a decreasing trend across the three conditions for post-completion errors, there was no reliable evidence to confirm the difference.

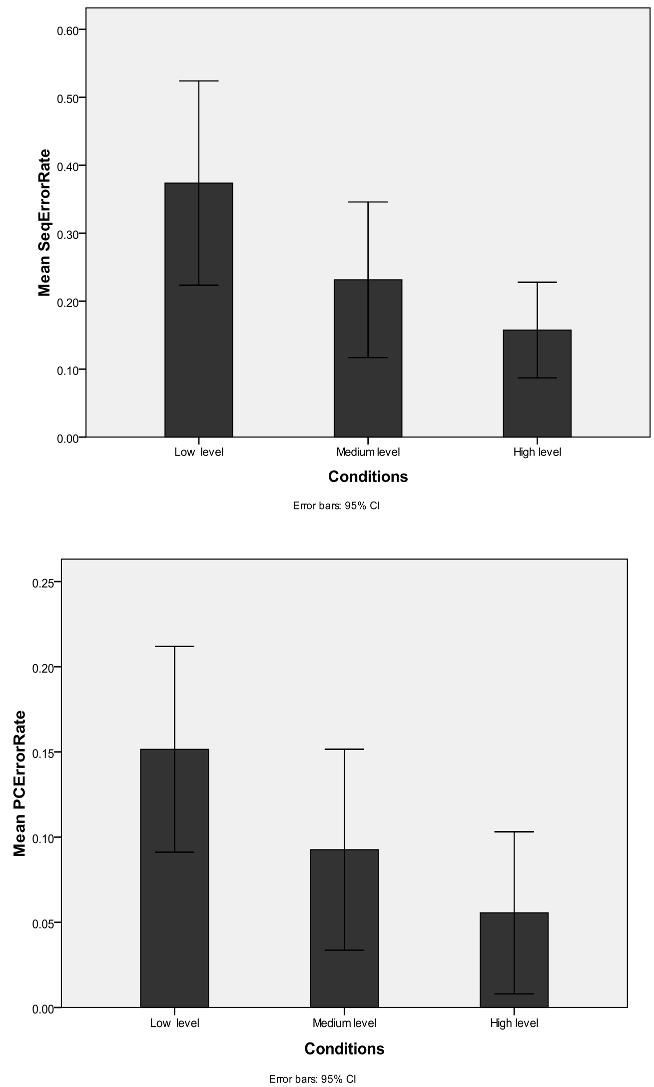


Figure3. Mean error rates of sub-task sequence errors and post-completion errors across the three conditions

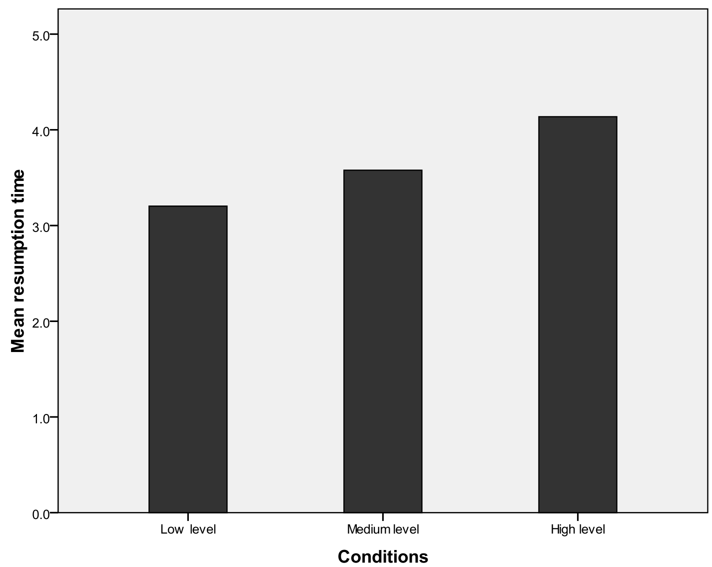
Task resumption lag

Figure4 shows the contrast of mean resumption lag across the three conditions. The left bar-chart in Figure4 reveals the mean overall resumption lag over low level, medium level and high level conditions. From the chart, we can tell the overall resumption lag presented a rising tendency. The mean length in low level condition was 3.20s (SD=2.01), and the mean duration in medium and high level conditions were 3.58s (SD=2.69) and 4.13s (SD=3.19), respectively. An ANOVA was introduced to verify the difference observed over the bar-chart ($F(2, 32) = 6.35$, $p<0.05$). Then, the results of Tukey HSD post-hoc comparison showed revealed that the mean resumption lag was different significantly among low level versus medium level; low level versus high level and medium level versus high level (all p's<0.05) The results enhance our observation from the bar-chart, which resumption lag made a increasing trend significantly across the three conditions.

To be more specific, the clustered bar-chart in the right part of Figure4 indicated the mean resumption lag when interruptions occurred immediately after one of the sub-tasks and just before post-completion step, respectively. The blue bars stood for the mean resumption lag the moment interruptions occurred immediately after one of the sub tasks, across which resumption lag increased orderly. The resumption lag in low level condition was 4.12s (SD=2.24). In medium and high level conditions, the lags were 4.61s (SD=3.20) and 5.38s (3.72), respectively. An ANOVA was used to identify the significance of effect on resumption ($F(2, 32) = 4.26$, $p<0.05$), suggesting that resumption lag specified for sub-task sequence errors had a reliable difference across the three conditions. To be more precise, the Tukey HSD post-hoc comparison was conducted for each combination between two out three conditions. There was significant difference between each combination between two out three conditions (all p's<0.05).

The green bars described the variation of mean resumption lag when interruptions occurred just before post-completion step across the three conditions. The resumption lag presented an ascending tendency as well. In the low level condition, the mean resumption lag was 2.28s (SD=1.18); the mean resumption lag in medium and high level conditions were 2.55s (SD=1.48) and 2.89s (SD=1.86), respectively. In light of the results of ANOVA ($F(2, 32) = 4.09$, $p<0.05$), there was significant effect on resumption lag across the three conditions. Then, in order to confirm the difference between every two conditions, a Tukey HSD post-hoc comparison was calculated (all p's<0.05).

Comparing to the resumption lag between recovering from interruptions that occurred immediately after one of the sub-tasks and just before post-completion step, it was found that recovering from interruptions that occurred immediately after one of the sub-task was more time-consuming than recovering from interruptions that occurred just before post-completion step.



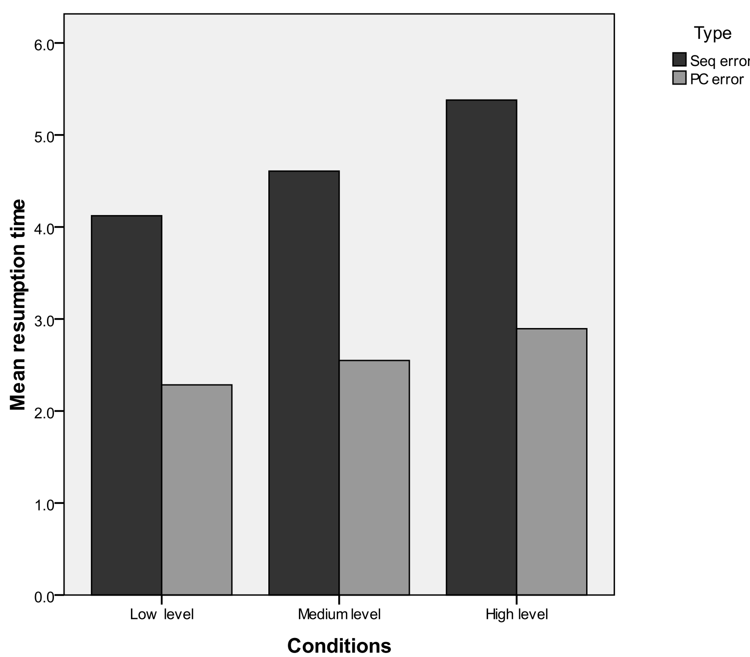


Figure4. Mean resumption lag across the three conditions

Relationships between error frequency and resumption lag

Figure5 illustrates the relationships between error frequency and resumption lag. From Number1 to Number3, the three plot points indicate the relationship between how fast participants conducted interruption recovery and how frequently sub-task sequence errors were made. From the figure, we can see task resumption was fastest in the low level condition, but the sub-task sequence error rate was highest (plot point No.1). Number 3 plot point lied in the most right side of the figure, which suggests that task resumption was slowest in the high level condition; meanwhile, the error-rate was lowest.

The three plot points in the left half of Figure5 present the relationships between the lengths of resumption lag and post-completion error rates across the three conditions. The most left plot point (No.4 plot point) shows time cost and post-completion error rate in low level condition, which means interruptions were recovered most quickly, but the errors were made most frequently. Number6 plot point shows the resumption lag and post-completion error rate in the high level condition, which is clear that task resumption in the high level condition was slowest, but the error rate was lowest.

Both sub-task sequence and post-completion errors presents a descending tendency in terms of error rates across the low level, medium level and high level condition; meanwhile, the resumption lag was in an ascending direction from low level condition to high level condition. In brief, Figure5 reveals that longer resumption lag could trade lower error rate in some extent.

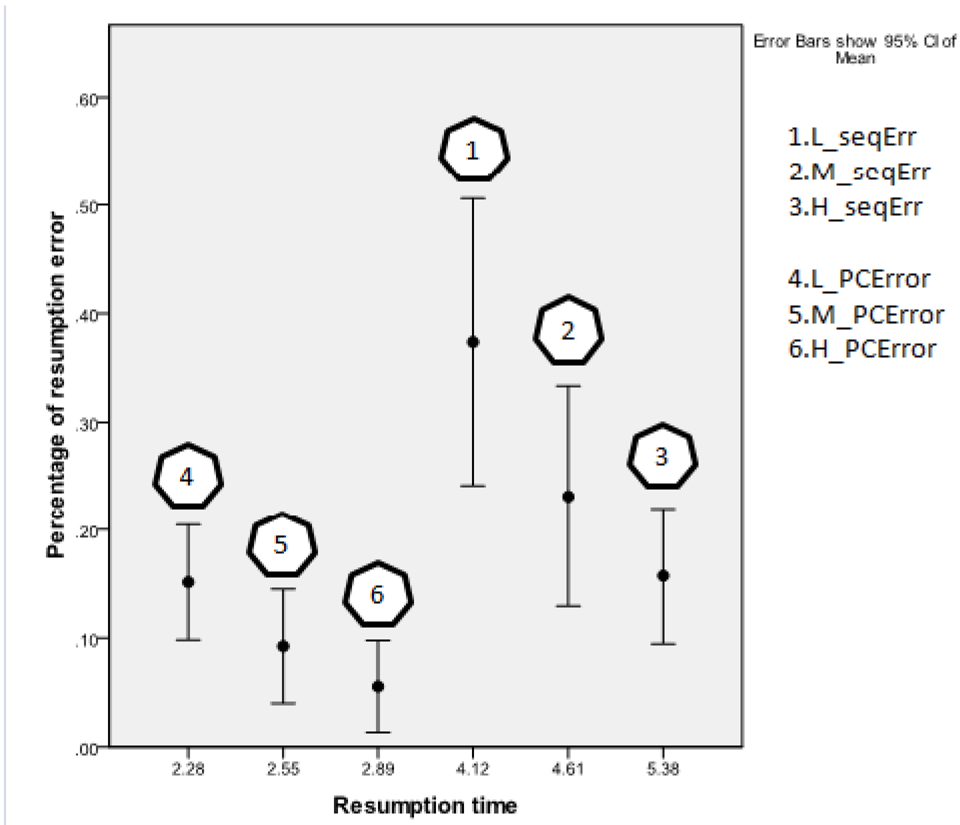


Figure5. The relationships between mean error rates and resumption lag

Figure6 describes the relationships between resumption lag and the number of sub-task sequence errors in another way, which aim to indicate the number of sub-task sequence errors in each specified resumption lag period. From the bar chart, we can see that the bars followed a declined trend except for the most right bar (over 8.5s). That is to say, with the increasing length of resumption lag, fewer sub-task resumption errors were made. When resumption lag was above 8.5s, the number of errors increased dramatically.

In sum, more errors occurred in both shortest and longest duration of resumption lag (0-2.5s and over 8.5s). In the middle (2.5s-4.5s, 4.5s-6.5s and 6.5s-8.5s), fewer errors were made.

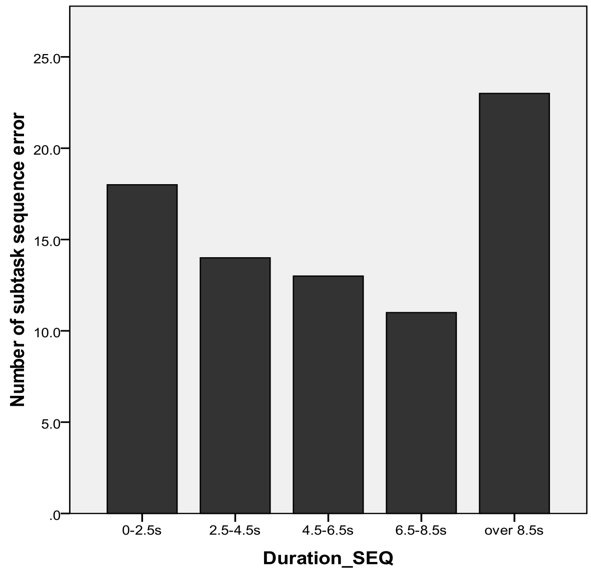


Figure6. The distribution of sub-task sequence errors in each specified time-period

Figure7 demonstrates the relationships between resumption lag and number of errors for post-completion errors. Overall, the number of post-completion errors was in a descending way with the increasing length of resumption lag. Most of the post-completion errors were made in the duration of 0-2.5s and 2.5-4.5s. From 4.5-6.5s, the number of post-completion errors reduced remarkably. Moreover, very fewer post-completion errors were made when resumption lag was above 8.5s.

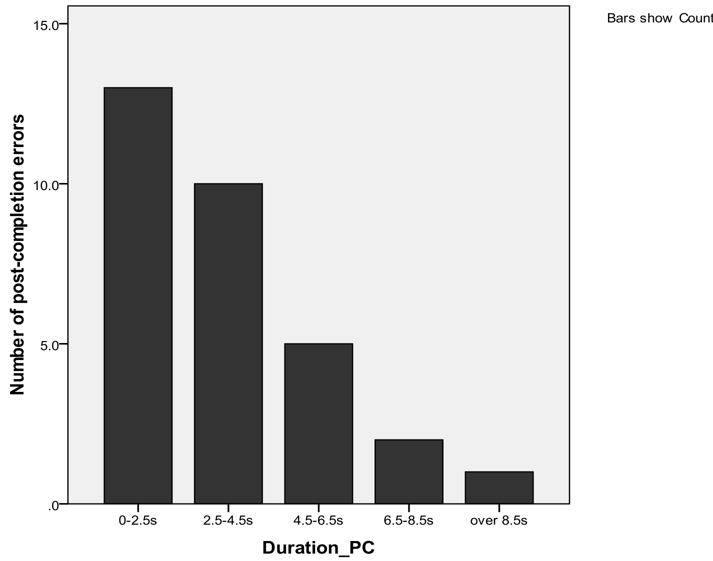


Figure7. The distribution of post-completion errors in each specified time-period

CHAPTER 4. GENERAL DISCUSSION

The experiment set different levels of time-costly penalty across the three conditions, which motivated people to reduce both sub-task sequence and post-completion errors when they recovered from interruptions. The aim of the experiment was to identify how people adjusted their strategies under the motivation to avoid penalty as much as they could.

According to the results, one major finding was related to error rates. On one hand, participants made fewer sub-task sequence errors orderly from the low level condition to the high level condition. The statistics indicated that there were reliable different error rates between the low level condition and the medium condition, as well as between the low level condition and the high level condition. Although the mean error rate in the high level condition was lower than the error rate in the medium level condition, the statistic analysis did not show significant difference. On the other hand, the mean error rates of post-completion errors across the three conditions also presented a descending tendency. However, the significance of difference was not reliable in light of statistic analysis, so we cannot conclude that participants reduced post-completion errors across the three conditions with their strategic controls.

The other finding in this experiment was in relation of resumption lag. Overall, across the three conditions, participants took longer resumption lag to conduct task resumption orderly from the low level condition to high level condition. Moreover, the statistics confirmed the reliability of difference. To be more specific, when interruptions occurred immediately after one of the sub tasks, participant took longest resumption lag to resume the primary task in the high level condition. Conversely, shortest period of resumption lag was observed when people recovered from interruptions in the low level condition. The statistic analysis confirmed the significant difference. When interruptions triggered just before post-completion step, as mention above, similar contrasts across the three conditions were detected. People spent most time to resume the primary task in the high level condition, and least time in the low level condition. Moreover, the statistics also revealed the difference was reliable. Compared with the length of resumption lag between recovering from interruptions that occurred immediately after one of the sub-tasks and just before post-completion step, it was found that people took longer time to resume the primary task when interruptions triggered immediately after one of the sub-task.

Tracing back to the hypotheses mentioned in the method chapter, for the sub-task sequence errors, the error-rates reduced significantly in the medium and high level condition compared with low level condition, which is consistent with our original prediction. About the conflicting prediction that whether or not sub-task sequence errors can be further mitigated from the medium level condition to the high level condition, the results showed that errors were not reduced further. In terms of post-completion errors, our findings are contradictive with our prediction. There was no reliable difference in error-rates across the three conditions. The other hypothesis is about the variation of resumption lag across the three conditions. The results are evident to match our initial prediction.

4.1. Relationships between sub-task sequence error rates and resumption lag

In terms of sub-task sequence errors, it was found that fewer sub-task sequence errors were made in the medium and high level conditions comparing to the low level condition. Meanwhile, the resumption lag extended from the low level condition to the high level condition orderly.

On one hand, the findings suggest that people can be motivated to mitigate errors by via strategic controls. This is consistent with the AGM (Altmann & Trafton, 2002) model, in which it is proposed task resumption can be benefit from active rehearsals and memory retrieval before or during the interruptions. Moreover, Moreover, the results matched Cades et al.'s study (2007) as well. In the study, they proposed that during interruptions, rehearsals were crucial to facilitate and improve task resumption. In the medium and high level conditions of our experiment, due to the costly penalty followed by resumption errors, participants tended to rehearse the suspended goals and strengthen the activation of suspended goals before or during the duration of interruptions in order to avoid the costly penalty. Consequently, the error rates were reduced dramatically. Our findings also enhanced Monk, Trafton and Boehm-Davis (2008)'s empirical study in some extent, in which participants were required to conduct a sequential procedural with occasional interruptions. The error rate in the condition that set an 8-second period of interruption lag was significantly lower than the condition without interruption lag.

On the other hand, although people made fewer errors in the medium and high level conditions, they took longer time to resume the primary task. This seems to be contradictive with the AGM (Altmann & Trafton, 2002) model. In light of the AGM model, interruptions lead to memory decay for the goals related to the primary task. Moreover, memory decays continuously over time. Therefore, it is reasonable to conclude that the memory for suspended goals will decay further with longer resumption lag. Thus, more errors should be made in the medium and high level conditions.

In one way, the contradiction can be explained by a piece of study conducted by Monk, Trafton and Boehm-Davis (2008). In the study, Monk et al. enhance the assumption proposed in the AGM model, which memory for suspended goals does decay with increasing time. Meanwhile, they explained why some early studies (e.g., Gillie & Broadbent, 1989) were failed to identify the relationships between interruption length and resumption lag. They predict the one possible cause is that when an interruption lasts over a certain period of time, the memory decay approaches an asymptote. That is to say further extended resumption lag does not affect activation level of suspended goals significantly. Although Monk et al. did not suggest explicit interruption length such that memory level tends to be asymptotical; we predict that a 30-second period of interruption set in our experiment is sufficiently long to eliminate the significant difference of memory level when participants took longer resumption lag in the medium and high level condition than the resumption lag taken in the low level condition. Therefore, participants in the medium and high level conditions made fewer errors due to their active rehearsal strategy. The relatively long resumption lag did not reduce suspended goals activation level significantly.

In the other way, the contradiction can be explained by the AGM (Altmann & Trafton, 2002) model as well. Although memory decays with increasing time, the rehearsals can strengthen activation of memory. Therefore, the memory for suspended goals in the medium and high level conditions did not decay continuously. Participants might conduct actively multiple rehearsals both before and during the interruptions to prevent memory decay. Similar demonstration is also mentioned in Badeley's study (2000), which proposed that memory can be retrieved and strengthened via rehearsals. Therefore, comparing to the resumption lag in low level condition, the longer resumption lag in medium and high level conditions did not reduce the memory activation level effectively. Thus, fewer errors were made in the medium and high level conditions. Conversely, although participants in the low level condition spent less time to recover from the interruption, they might not conduct any rehearsals before or during the interruptions, and the memory for the suspended goals decayed persistently. After interruptions, they tended to guess the next action by ambiguous memory or instinct. Thus, more errors were observed in the low level condition.

However, in the medium and high level conditions, why participants consumed longer resumption lag to resume the primary task is still left a question. One possible explanation is the participants were strongly motivated to avoid costly penalty, so they tended to check their answer before making decision in order to avoid click on an incorrect button accidently. Another possibility might be due to indeterminacy of their answers. Although they conducted active rehearsals before or during the interruptions, they were not confident sufficiently to make decision immediately. Thus, they tended to analyze the contexts and at the same time to search for reliable evidence, and then made decision. A piece of study conducted by Lau, Joannidis and Schmid (1998) supports our prediction. The study proposed that it is necessary to sum up reliable and logical evidence before submitting an answer in order to improve accuracy.

About the relationships between medium and high level conditions, the findings indicated that there was no reliable difference on error rates, but the resumption lag in the high level condition was longer than the resumption lag in the medium level condition, which is inconsistent with the theory of speed-accuracy trade-offs theory. But, the results can be explained by the AGM (Altmann & Trafton, 2002) model, which demonstrated that previous goals are stored at an "interference threshold" which means the memory for suspended goals can be retrieved only the moment their activation level is beyond the threshold. The length of time that people need to task to retrieve the suspended goals is closely related to the level of activation above threshold. In other words, if the activation level of suspended goals is under threshold, longer resumption lag is no use. In our study, compared with the medium level condition, it was reasonable to predict that participants spent longer time to recover from interruptions in the high level condition due to more costly penalty that results in more strong motivation to reduce resumption errors. However, the level of activation might be below the threshold value, so the longer resumption lag did not trade higher accuracy. Similar explanation is also mentioned in Reed's study (1973). An experiment is designed to identify the speed-accuracy trade-offs effects on cognition memory in the study. The results reveal that there was no further significant improvement on accuracy when resumption lag was over four seconds. In other words, the accuracy level presents as asymptote when response delay exceeds four second. Therefore, compared with medium level condition, further extended resumption lag in the high level condition did not trade higher accuracy.

In order to investigate into the relationships between error rates and resumption lag in a visual way. We calculated the number of sub-task sequence errors in each time period (0-2.5s; 2.5-4.5s; 4.5-6.5s; 6.5-8.5s; and over 8.5s). From the period of 0-2.5s to 6.5-8.5s, the number of errors reduced orderly, but during the period of over 8.5s, the number of errors increased remarkably. As mention above, we predict this is because the activation level did not reach threshold, memory could not be retrieved despite taking very long resumption lag.

4.2. Relationships between post-completion error rates and resumption lag

In light of the results in terms of post-completion errors, although there was reliable different resumption lag across the three conditions, there was marginal significance ($F(2, 32) = 2.74$, $p=0.08$) of effect on the error rates across the three conditions. Therefore, we are not confident to conclude the speed-accuracy trade-offs effect on post-completion errors. This is consistent with some studies (e.g., Back et al., 2008; Byrne & Bovair, 1997). For example, Back et al.'s study (2008) adopted two gaming settings to design their experiments, in which participants were motivated to avoid post-completion errors. In the experiments, once participants made post-completion errors, their marks would be cleared. The results showed that post-completion errors were not reduced significantly. Some other studies (e.g., Ratwani, McCurry & Trafton's study, 2008; Barne & Bovair, 1997; Chung & Byrne, 2007; Li et al., 2005) suggest that post-completion errors can be mitigated in three ways: redesigning the system, offering visual cues, and predicting possibility of potential errors.

In summary, this study is interested in identifying whether or not people can be motivated to mitigate both sub-task sequence and post-completion errors by changing strategy during task resumption processes. Based on the findings in our experiment, we discussed why people made fewer sub-task sequence errors in the medium and high level conditions and at the same time, took longer resumption lag to recover from the interruptions compared with the situations in the low level condition. For the post-completion errors, we did not conclude the significant difference across the three conditions, which is consistent with previous work. Subsequently, we plan to discuss the limitations in this experiment followed by recommendation for future work in this area. Finally, we will consider the implication of this study.

4.3. Limitations

The experiment was design to identify how people adjust their resuming strategies to reduce both sub-task sequence errors and post-completion errors during task resumption. A couple of tasks adapted from Li et al.'s (2008) Doughnut machine task and Packing task were introduced into our experiment. Two issues need to be improved are shown as follows.

Firstly, the post-completion step was "Archive" button in this experiment. Participants clicked on the "Process" button to submit a pharmacy order, and then the button label changed to "Archive" automatically. After clicking "Archive" button, the label changed back to "Process" once again. That is to say, only after the "Process" button was clicked, the "Archive" button could be shown in the task screen. If participants detected this defect in interface design, they could recover from the interruptions that occurred just before post-completion step easily. According to the results, the error-rates of post-completion errors were much lower than the error rates of sub-task sequence errors in each condition. Moreover, we did not conclude significant distinction of post-completion errors across the three conditions. Although the consequence mentioned above cannot attribute to interface design completely, there may be some bias for the data collected before post-completion step. Therefore, we suggest that a separate button can be set as post-completion step in the task screen.

Secondly, as mention above, our experiment was adapted from previous work (Li et al., 2008), but our tasks were much simpler than previous tasks. About the main task (Pharmacy-ordering task), the number of sub tasks was reduced to three, and there was only one interruption occurred in each trial. For each test session, participants needed to complete five training trials and 27 formal trials. We predict that participants tend to be impatient to conduct relatively monotone task repeatedly, which may have negative effects on their performance. For example, if participants felt boring to execute the task in the later period of the test session, they may not be motivated sufficiently to rehearse the suspended goals actively before or during the interruptions.

Several issues that limited the performance of this experiment were due to my inappropriate operation and external environment. There was no detailed document that was used to explain the tasks provided to participants. Before the experiment started formally, I showed an operation manual diagram to participants and at the same time describe the tasks verbally. It is not realistic to describe the tasks with completely same words. The experiment lasted approximately 45 minutes, and participants were paid £3 for their time, which was significantly below the rate that UCL pays participants. During this summer, UCL Bedford way building is conduct construction work. The noise may lead to distraction.

4.4. Recommendation and future experimentation

Several potential avenues for further study in this field are proposed in this section.

Firstly, about post-completion errors, our experiment did not find sufficient evidence that showed people could be motivated to mitigate post-completion errors. Similarly, a piece of study does not conclude significant effect of motivation on post-completion error rates (Back et al., 2008). Many studies focus on reducing post-completion errors with external cues, system redesign and advanced prediction (e.g., Barne & Bovair, 1997; Chung & Byrne, 2007; Li et al., 2005; Ratwani, McCurry & Trafton, 2008). Byrne and Bovair (1997) explain that post-completion errors are due to high working memory load. Currently, there are no effective studies that interpret how people can be motivated to reduce this type of error. Therefore, in order to investigate this issue further, we propose that it may be appropriate to design experiment associated with Byrne and Bovair’s model (1997). To be more specific, the main reason for post-completion errors is due to working memory load (Barne & Bovair, 1997), therefore, it is necessary to identify how to motivate people to reduce working memory load. Mayer and Moreno (2003) suggest a number of ways, such as time allowance, to mitigate cognitive overload, which may be valuable for future experiment. Therefore, maybe, we can set a period of time for releasing working memory load after interruptions that occurred just before post-completion step. In this way, people can make speed-accuracy trade-offs during the time allowance. Back, Brumby and Cox set an enforced 10-second period of locked-out time after interruptions in their experiment. And the results show that it is effective to reduce sequence errors. We wonder what happens to post-completion errors.

In terms of sub-task sequence errors, participants made speed-accuracy trade-offs between the low level condition and either the medium or high level condition in our experiment. But there was no significant speed-accuracy trade-offs effect between the medium and high conditions. Therefore, we predict that there may be an optimal range of resumption lag during task resumption. Reed (1973) also predicts the optimal response time in order to improve cognition memory in his study. In the future, we plan to conduct an experiment with a set of intervals (i.e., 3s, 6s and 9s) after interruption to identify the difference of task resumption performance across various resumption lag.

4.5. Implications

The results of this study indicated that people can be motivated to make speed-accuracy trade-offs in terms of sub-sequence errors during task resumption processes. Our findings would be meaningful and invaluable for the workspaces where interruptions occur frequently.

In some busy workspaces, such as financial organizations and commercial environments, interruptions occur very frequently (Sullivan et al., 1996; Sullivan, Vardell & Johnson, 1997). The findings of our study can encourage people to slow down after an interruption in order to improve accuracy.

In the domains with high request of reliability and safety, such as healthcare organisations, resumption errors because of interruptions may result in serious consequence (Back, Brumby & Cox, 2010; Chisholm et al., 2000; Li et al., 2006; Tucker & Spear, 2006). Our findings suggest that speed-accuracy trade-offs theories can be applied to design medical devices which assist nurses in reducing medical accidents.

Our study did not conclude that people can be motivated to mitigate post-completion errors, but this encourages researchers to search for significantly effective way to alleviate this type of error in future study. Most previous studies focuses on studying how to reduce post-completion errors with extra assistance, such as external cues (e.g., Chung & Byrne, 2004); or how to redesign the current system in move out post-completion step (e.g., Byrne & Bovair, 1997; Li et al., 2005). Our study attempted to reduce post-completion errors with an economical and easy way, which may arouse more attention in the field of HCI to investigate post-completion errors.

CHAPTER 5. CONCLUSION

This study aims to identify whether or not people can be motivated to make speed-accuracy trade-offs on both sub-task sequence and post-completion errors in a procedural routine followed by an interruption. The experiment involved a primary task (Pharmacy-ordering task) in which participants were required to complete multiple prescription sheets in correct sequence, and an interrupting task (Packing task) which was relevant mental arithmetical problems. During the execution of the Pharmacy-ordering task, participants were interrupted by the Packing task immediately after one of the sub-tasks or just before post-completion step. Three conditions including the low level condition, medium condition and high level condition were defined by varied level of time-cost penalty after a resumption error.

In terms of sub-task sequence errors, it was found that the mean error-rate in the low level condition was significant higher than the error rates made by both the medium and high level conditions. Compared with the error-rate in medium level condition, the error-rate made by the high level condition did not decrease reliably. In terms of post-completion errors, participants made fewer errors orderly across the three conditions from low level to high level. However, the significance of difference across the three conditions was marginal.

Regardless of the interruptions occurred immediately after one of the sub-tasks or just before post-completion step, the length of resumption lag was significantly different across the three conditions. Participants in the high level condition took longest resumption lag to resume the primary task, and participants in the low level condition recovered from the interruption most quickly.

It is verifiable that people can be motivated to trade speed for fewer sub-task sequence errors. But the situation of post-completion errors left unclear. About the extending of current study, a couple of further research issues are proposed for future experiment: Reducing working memory load to mitigate post-completion errors; identifying the optimum duration of resumption lag.

These findings are meaningful in terms of device design to motivate people make speed-accuracy trade-offs.

REFERENCES

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

Altmann, E. M. & Trafton, J. G. (2004). Task interruption: Resumption lag and the role of cues. In *Proceedings of the 26th annual conference of the Cognitive Science Society* (pp. 42– 47). Hillsdale, NJ: Erlbaum.

Altmann, E. M., & Trafton, J. G. (2007). Timecourse of recovery from task interruption: Data and a model. *Psychonomic Bulletin & Review, 14*, 1079-1084.

Anderson, J. R. & Lebiere, C. (1998). *Atomic components of thought*. Mahwah, NJ: Erlbaum.

Back, J., Brumby, D. and Cox, A. (2010) Locked-out: investigating the effectiveness of system lockouts to reduce errors in routine tasks. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, New York, 2010, 3775–3780.

Back, J., Cheng, W.L., Dann, R., Curzon, P., Blandford, A. (2006). Does being motivated to avoid procedural errors influence their systematicity? *Proc. HCI 2006,* pp.151-157.

Baddeley, A. D. (2000) Working memory: The interface between memory and cognition. In: M. Gazzaniga (Ed.) *Cognitive Neuroscience: A Reader* (pp. 292-304). Oxford, UK: Blackwell Publishers Ltd.

Blandford, A. (2000) Designing to avoid post-completion errors. PUMA working paper http://www.eis.mdx.ac.uk/puma/wp33.pdf.

Boldini, A, Russo, R, & Avons, S.E. (2004). One process is not enough! A speed-accuracy tradeoff study of recognition memory. *Psychonomic Bulletin & Review, 11*, 353-361.

Byrne, M. D., & Bovair, S. (1997). A working memory model of a common procedural error. *Cognitive Science, 21*, 31-61.

Cades, D. M., Trafton, J. G., Boehm-Davis, D. A., & Monk, C. A. (2007). Does the difficulty of an interruption affect our ability to resume? *Proceedings of the 51st Annual Meeting of the Human Factors and Ergonomics Society* (pp. 234-238). Santa Monica, CA: HFES.

Cellier, J., & Eyrolle, H. (1992). Interference between switched tasks. *Ergonomics*, *35(1)*, 25-36.

Chisholm, C. D., Collison, E. K., Nelson, D. R., & Cordell W. H. (2000). Emergency department workplace interruptions: Are emergency physicians “interrupt-driven” and “multitasking”? *Academic Emergency Medicine, 7 (11)*, 1239-1243.

Chung, P. H., & Byrne, M. D. (2004). Visual cues to reduce errors in a routine procedural task. *Proceedings of the Twenty-Sixth Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Chung, P. H., & Byrne, M. D. (2008). Cue effectiveness in mitigating postcompletion errors in a routine procedural task. *International Journal of HumanComputer Studies, 66,* 217–232.

Czerwinski, M., Cutrell, E., & Horvitz, E. (2000). Instant messaging: Effects of relevance and timing. In S. Turner & P. Turner (Eds.), *People and Computers XIV: Proceedings of HCI* (pp. 71–76). London, UK. British Computing Society.

Czerwinski, M., Horvitz, E., & Wilhite, S. (2004). A diary study of task switching and interruptions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '04* (pp. 175-182). New York, NY: ACM Press.

Edwards, M. B., & Gronlund, S. D. (1998). Task interruption and its effects on memory. *Memory, 6*, 665-687.

Forster, J., Higgins, E.T., & Bianco, A.T. (2003). Speed/accuracy decisions in task performance: Built-in trade-off or separate strategic concerns? *Organizational Behavior and Human Decision Processes 90 (2003)* 148–164.

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

Gloria Mark, Daniela Gudith, Ulrich Klocke: The cost of interrupted work: more speed and stress. *CHI 2008*: 107-110.

González, V. M., & Mark, G. (2004). "Constant, constant, multi-tasking craziness:" Managing multiple working spheres. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 113-120). Vienna, Austria: ACM Press.

Hick, W. E., 1952. On the rate of gain of information. *Quarterly Journal of Expcrimcntal Psychology 4*, 1 l-26.

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 (1)*, 103-115.

Bailey, B.P., Konstan, J.A. and Carlis, J.V. (2000). Measuring the Effects of Interruptions on Task Pedrformance in the User interface. *IEEE Conference on Systems, Man and Cybernetics. Nashville*, TN.

Lau, J., Ioannidis, J. P., & Schmid, C. H. (1998). Summing up evidence: one answer is not always enough. *Lancet* 1998; **351:** 123–27.

Li, S. Y. W., Blandford, A., Cairns, P. and Young, R. M. (2005) Post-completion errors in problem solving. *27th Annual Conference of the Cognitive Science Society,* 1278–1283 29,30. Li et al., 2006, 2008;

Li, S. Y. W., Blandford, A., Cairns, P., & Young, R. M. (2008). The effect of interruptions on postcompletion and other procedural errors: An account based on the activation-based goal memory model. *Journal of Experimental Psychology: Applied, 14*, 314-328.

Li, S. Y. W., Cox, A. L., Blandford, A., Cairns, P., Young, R. M., & Abeles, A. (2006). Further investigations into post-completion error: the effects of interruption position and duration. In *Proceedings of the 28th Annual Meeting of the Cognitive Science Conference* (pp. 471-476). Vancouver, BC: Cognitive Science Society.

Mark G, Gudith D, Klocke U. The cost of interrupted work: more speed and stress. *In: ComputereHuman interactions (CHI) Florence, Italy*: ACM, 2008:107e10.

Mark G., Gudith D., Klocke U. (2008) The cost of interrupted work: More speed and stress, in: *Proceeding of the Twenty-Sixth Annual SIGCHI Conference on Human Factors in Computing Systems, CHI'08*, New York: ACM, 107-110.

McFarlane, D. C. (2002). Comparison of four primary methods for coordinating the interruption of people in human–computer interaction. *Human–Computer Interaction, 17,* 63–139.

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

Monk C. A., Boehm-Davis D. A., & Trafton J. G. (2004). Recovering from interruptions: Implications for driver distraction research, *Human Factors, 46 (4)*, 650-663.

Monk C. A., Boehm-Davis D. A. & Trafton J. G. (2004) Very brief interruptions result in resumption cost. *Proceedings of the Twenty-Sixth Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Monk, C. A., Trafton, J. G., & Boehm-Davis, D. A. (2008). The Effect of Interruption Duration and Demand on Resuming Suspended Goals. *Journal of Experimental Psychology: Applied, 14(4)*, 299-313.

Norman, D. A. (1981). Categorization of action slips. *Psychological Review, 88(1),* 1-15.

Norman, D. (1990). The design of everyday things. Doubleday, New York.

Oulasvirta, A., & Saariluoma, P. (2004). Long-term working memory and interrupting messages in human-computer interaction. *Behaviour & Information Technology, 23(1)*, 53-64.

Peterson, L. R., & Peterson, M. J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology, 58*, 193-98.

Prevost, M.-C., Aubin, F., Gauthier, J., & Mailhot, R. (2007). Preventing human errors in power grid management systems through user-interface redesign. *Systems, Man and Cybernetics, 2007. ISIC. IEEE International Conference*, 626 – 631.

Ratwani, R. M., McCurry, J. M., & Trafton, J. G. (2008). Predicting Postcompletion Errors using Eye Movements. In *Proceeding of the TwentySixth Annual SIGCHI Conference on Human Factors in Computing Systems, CHI '08* (pp. 539-542). New York, NY: ACM Press.

Reason, J. T. (1984). Lapses of attention in everyday life. In R. Parasuraman & D. R. Davies (Eds.), *Varieties of attention.* New York, NY: Academic Press.

Reason, J., 2002. Combating omission errors through task analysis and good reminders. *Quality and Safety in Healthcare* 11, 40–44.

Reed, A. V. (1973). Speed-accuracy trade-off in recognition memory. *Science, 181*, 574-576.

Speier C., Valacich J. S., & Vessey I. (1999). The influence of task interruption on individual decision making: An information overload perspective, *Decision Sciences, 30 (2)*, 337-360.

Sullivan, M. J., Suddeth, B. N., Vardell, T., & Vojdani, A. (1996). Interruption Costs, Customer Satisfaction and Expectations for Service Reliability. *IEEE Transactions on Power Systems,* Vol. 11, ‘No. 2, May 1996.

Sullivan, M. J., *Member*, *IEEE*, Vardell, T., & Johnson, M. (1997). Power Interruption Costs to Industrial and Commercial Consumers of Electricity. *IEEE TRANSACTIONS ON INDUSTRY APPLICATIONS,* VOL. 33, NO. 6, NOVEMBER/DECEMBER 1997.

Trafton, J. G., Altmann, E. M., Brock, D. P., & Mintz, F. (2003). Preparing to resume an interrupted task: effects of prospective goal encoding and retrospective rehearsal. *International Journal of Human Computer Studies*, *58(5)*, 583-603.

Trafton,J. G., Altmann, E. M., & Brock, D. P. (2005). Huh,what was I doing? How people use environmental cues after an interruption. In *Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting* (pp. 468–472). Santa Monica, CA: Human Factors and Ergonomics Society.

Trafton, J. G., Altmann, E. M., & Ratwani, R. M. (2011). A memory for goals model of sequence errors. *Cognitive Systems Research 12 (2011),* 134–143.

Tucker, A. L., & Spear, S. J. (2006). Operational failures and interruptions in hospital nursing. *Health Services Research, 41,* 643–662.

Wickelgren, W. A. (1977). Speed-Accuracy Tradeoff and Information Processing Dynamics. *Acta Psychologica, 41,* (1977) 67-85.

Woodworth, R. S. (1899). Accuracy of voluntary movement. *Psychological Review, 3,* 1-106.

Zijlstra, F. R. H., Roe, R. A., Leonova, A. B., & Krediet, I. (1999). Temporal factors in mental work: Effects of interrupted activities. *Journal of Occupational and Organizational Psychology*, *72(2)*, 164–185.

APPENDIX - EXPERIMENT INTRODUCTION

Overview of this experiment

Overall, this experiment is trying to identify how people complete a procedural routine. The whole experiment lasts for approximately one hour. The brief agenda of completed experiment is shown as follows.

- Sign the consent form.
- Read experiment overview and operation manual.
- Training section (complete 5 trials).
- Formal experiment section (complete 27 trials).
- Break in the halfway (roughly 5 minutes).
- Sign the payment form and receipt.
- Debrief or feedback (optional).

In brief, the scenario of the task is about processing the prescription order. Participants need to fill in relevant information based on the prescription details. The task starts from “Next Prescription”, and followed by “Type”, “Shape”, “Colour”, “Process” and “Archive”. When an experimental trial begins, firstly, participants need to click the “Next prescription”; then the detailed prescription is displayed in the central of the system screen. Subsequently, the participants are required to click “Type” button and fill in the detailed prescription information in “Type” area, following with “Shape” and “Colour” in similar operation. When the prescription is filled completely, participants click “Process” to submit the order. Finally, “Archive” is a post-complete step, participants click this button in order to clean the current order and prepare for the next prescription order.

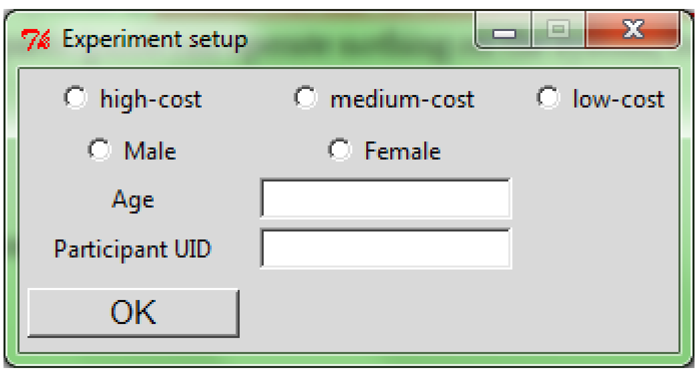
During the trials, you may be interrupted and required to complete a secondary task. This task requires you to calculate how many packs are required to pack each batch of medication. When the interrupting task finishes, you will be returned to the main task and should continue from where you left off.

If you restart the trial at the incorrect point following an interruption, the machine will break. You will then have to wait for a short period of time while the machine is being fixed.

Your aim is to complete 27 trials as quickly as possible.

Operation guide

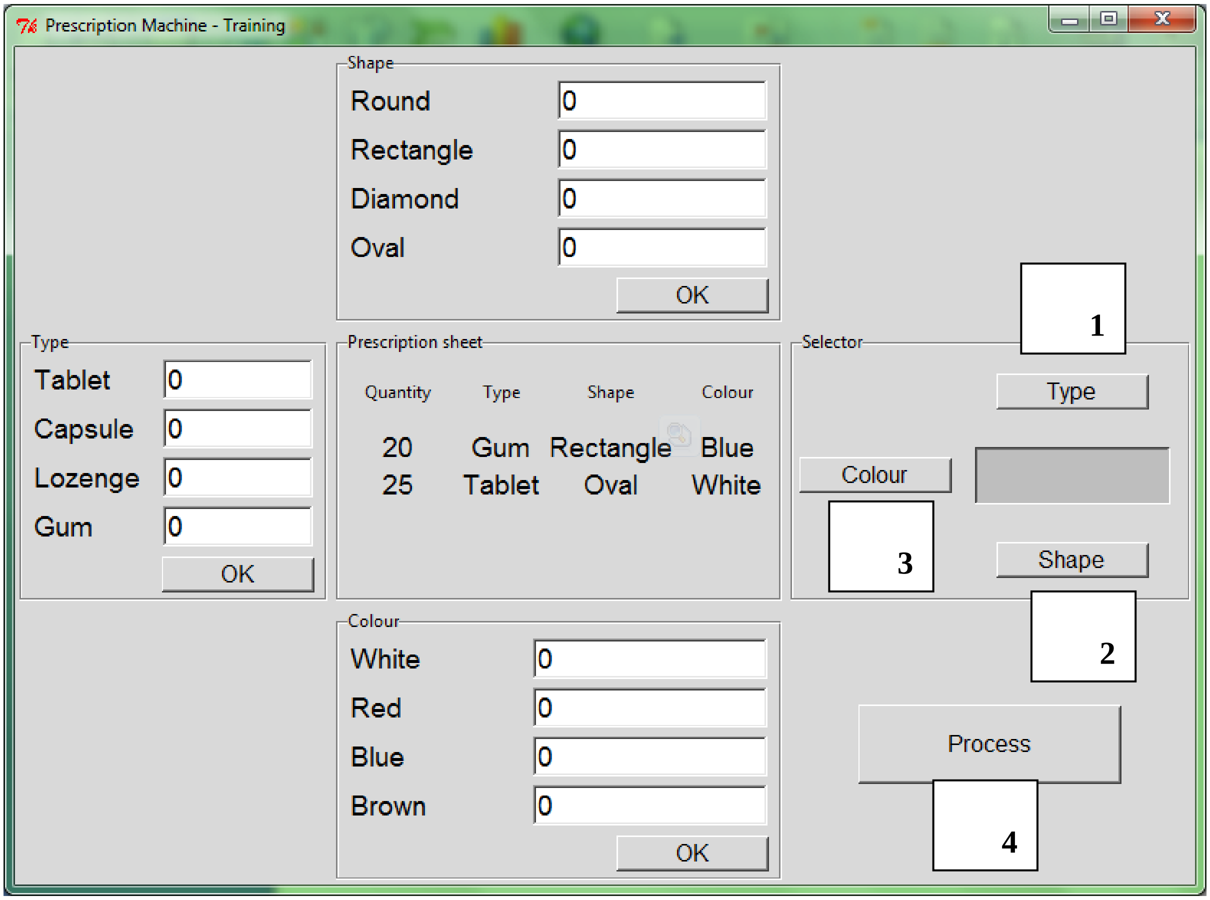
1. Input necessary personal information

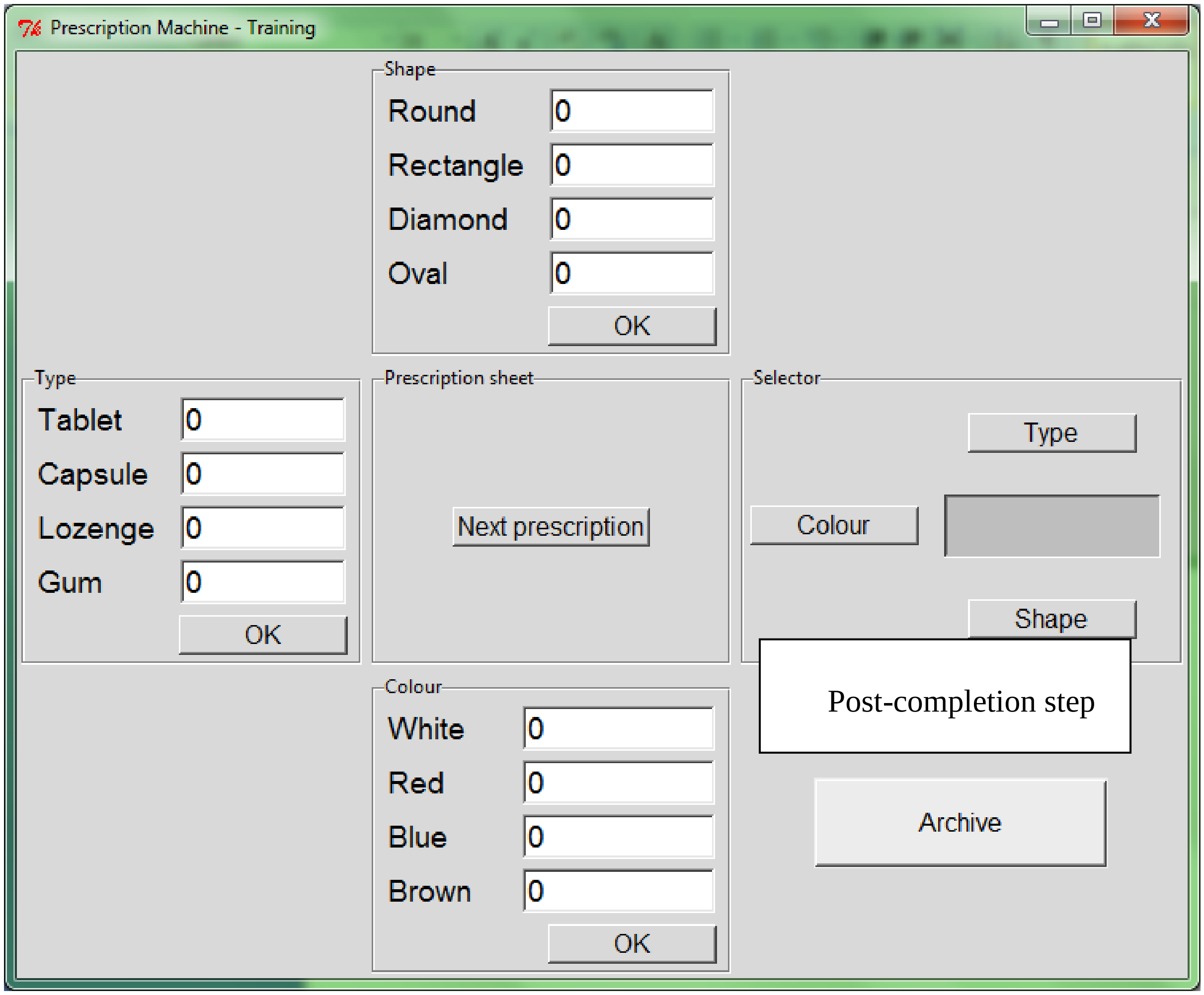


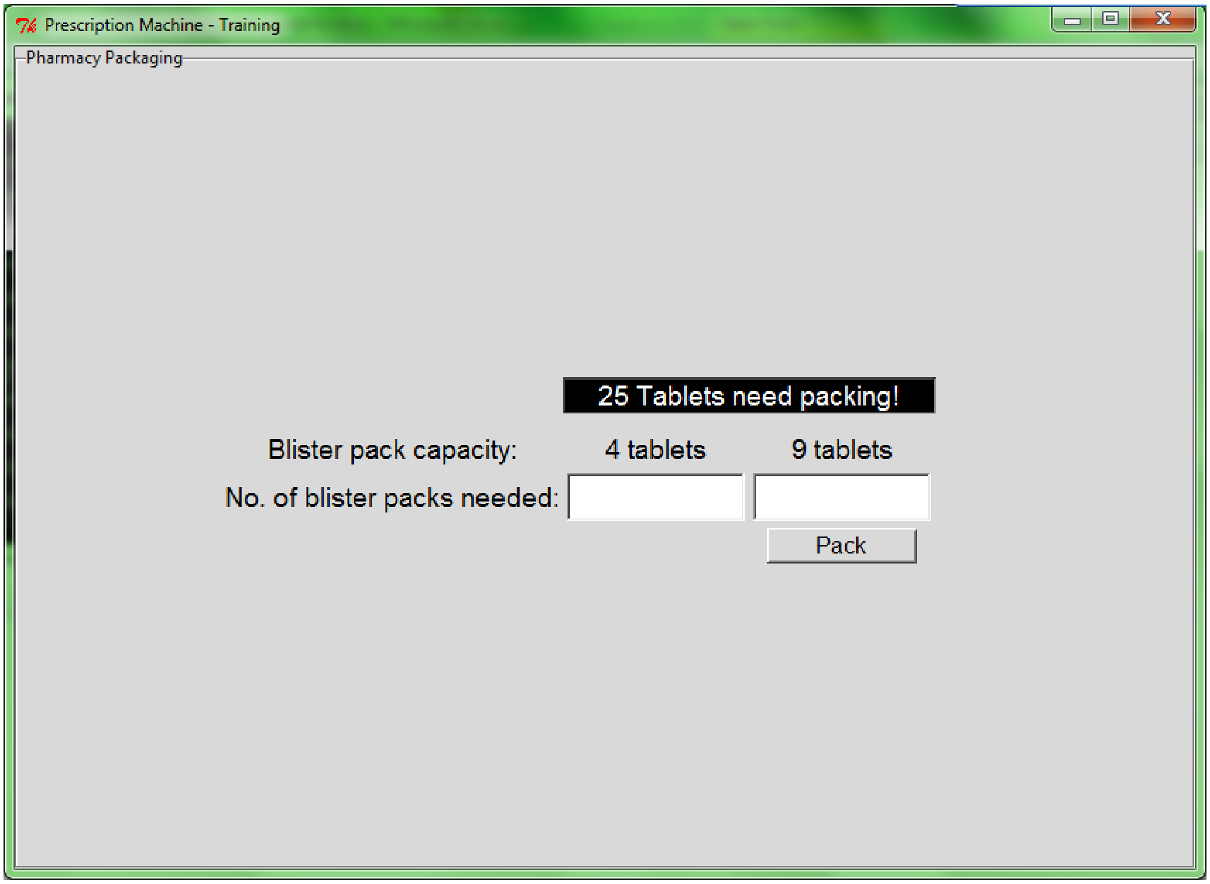
1.1 The experiment includes high-cost, medium-cost and low-condition. The participants will be dispatched one of them randomly.

1.2 After the participants sign the informed consent form, they will be allocated an participant user ID

2. Operation manual


Prescription Machine - Training
Shape
Round 0
Rectangle 0
Diamond 0
Oval 0
OK
Type
Tablet 0
Capsule 0
Lozenge 0
Gum 0
OK
Prescription sheet
Quantity Type Shape Colour
20 Gum Rectangle Blue
25 Tablet Oval White
Selector
Type
Colour
Shape
1
2
3
Colour
White 0
Red 0
Blue 0
Brown 0
OK
Process
4

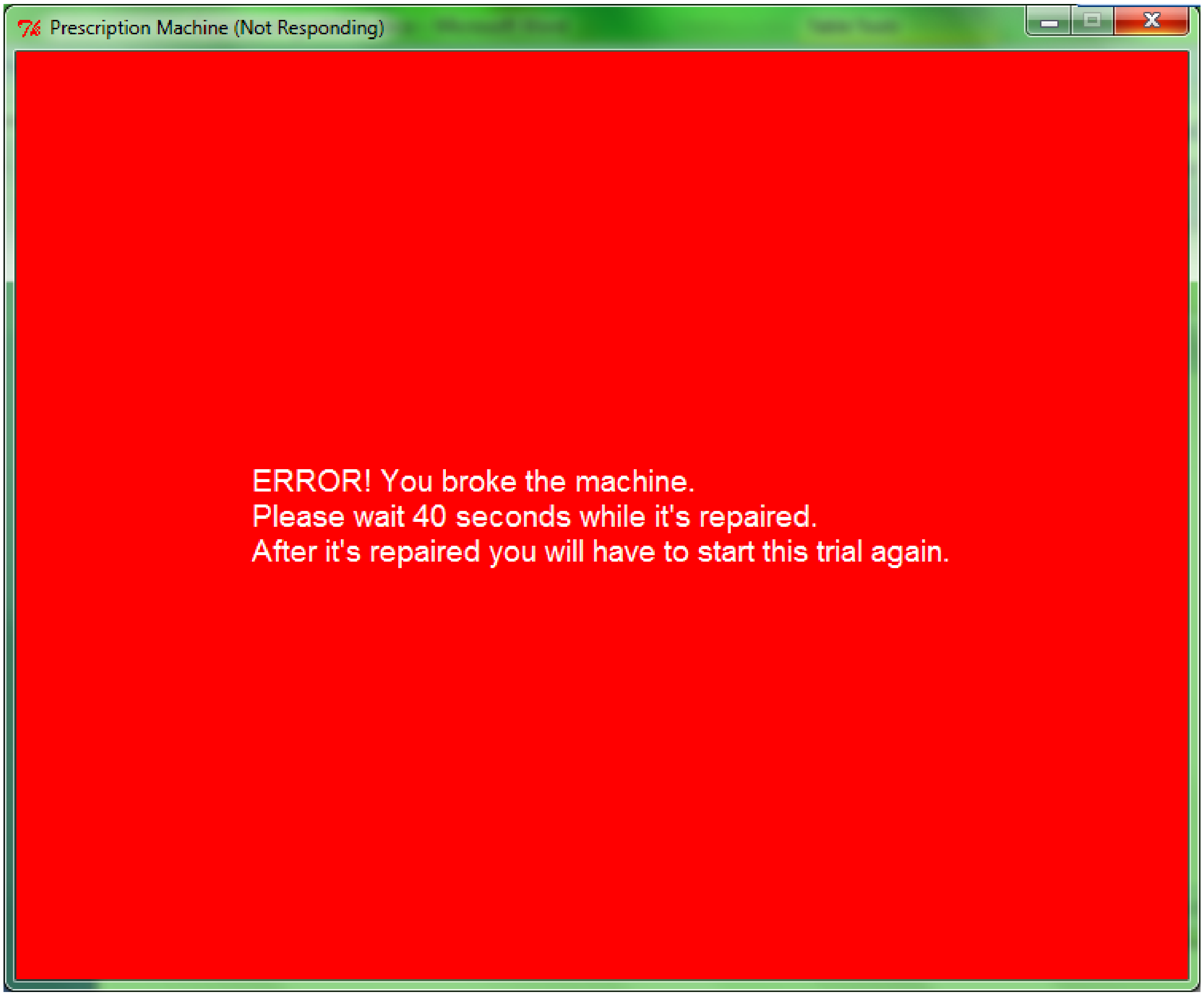

Prescription Machine - Training
Shape
Round
0
Rectangle
0
Diamond
0
Oval
0
OK
Type
Tablet
0
Capsule
0
Lozenge
0
Gum
0
OK
Prescription sheet
Next prescription
Selector
Type
Colour
Shape
Post-completion step
Archive
Colour
White
0
Red
0
Blue
0
Brown
0
OK



As a typical example shown above, this task is about using the 4 tablets and 9 tablets blister pack respectively to place 25 tablets. For instance,

$4 * 4 + 9 * 1 = 25.$

The lockout screen:



If the participants recover from the interruption at wrong subtask, the system will lock out for 40 seconds in high-cost condition, or 20 seconds in medium-cost condition. During the locked-out, the system has no available interface to be operated.