Public Paper no. 108

Recovering from an Interruption: Investigating Speed-accuracy Tradeoffs in Task Resumption Behavior

Duncan Brumby, Anna Cox, Jonathan Back & Sandy Gould


PP release date: 13 February 2013
Recovering from an Interruption: Investigating Speed-Accuracy Tradeoffs in Task Resumption Behavior

Duncan P. Brumby, Anna L. Cox, Jonathan Back, and Sandy J. J. Gould

University College London

Author note

Duncan P. Brumby, Anna L. Cox, Jonathan Back, and Sandy J. J. Gould are at the UCL Interaction Centre, University College London, United Kingdom.

This work was funded by EPSRC Grants EP/G059063/1 & EP/G043507/1. We thank Ann Blandford, Richard Young, Chris Janssen, Maartje Ament, Paul Curzon and the CHI+MED team for useful comments on this work. We also wish to thank Simon Li for use of his code for implementation of the tasks used here. This paper has been much improved by the suggestions from Wendy Rogers, Greg Trafton, David Cades, and an anonymous reviewer.

A preliminary report of Experiment 2 was presented as Work-in-Progress at the ACM SIGCHI conference, CHI 2010 (Back, Brumby, & Cox, 2010).

Correspondence concerning this article should be addressed to Duncan Brumby, UCL Interaction Centre, University College London, London WC1E 6BT, United Kingdom. E-mail: Brumby@cs.ucl.ac.uk
Abstract

Interruptions are disruptive because they take time to recover from, in the form of a resumption lag, and lead to an increase in the likelihood of errors being made. Despite an abundance of work investigating the effect of interruptions on routine task performance, little is known about whether there is a link between how quickly a task is resumed following an interruption (i.e., the duration of the post-interruption resumption lag) and the likelihood that an error is made. Two experiments are reported in which participants were interrupted by a cognitively demanding secondary mental arithmetic task while working on a routine sequential data-entry task. In Experiment 1 the time-cost of making an error on the primary task was varied between conditions. When errors were associated with a high time-cost penalty, participants made fewer errors and resumed the primary task more slowly than when errors were associated with a low time-cost penalty. In Experiment 2 participants were prohibited from resuming the primary task quickly by a 10-second system lockout period following the completion of the interrupting task. This lockout period led to a significant reduction in resumption errors because the lockout prohibited fast, inaccurate task resumptions. Taken together, our results suggest that longer resumption lags following an interruption are beneficial in terms of reducing the likelihood of errors being made. We discuss the practical implications of how systems might be designed to encourage more reflective task resumption behavior in situations where interruptions are commonplace.

Keywords: interruption, error, memory, speed-accuracy tradeoff
Recovering from an Interruption: Investigating Speed-Accuracy Tradeoffs in Task Resumption Behavior

Data-entry tasks are an everyday activity in a number of workplace settings. For instance, nurses set infusion pumps, office workers complete electronic forms, and pilots program flight paths. All of these kinds of tasks require the user to enter values into a device using a routine sequential procedure (Card, Moran, & Newell, 1983). They are normally executed easily and without error. However, because of the complex and demanding nature of many workplace settings, users are occasionally interrupted partway through a task and have to resume it later. Indeed, interruptions have been well documented in healthcare settings (e.g., Chisholm, Collison, Nelson, & Cordell, 2000; Grundgeiger, Sanderson, Venkatesh, & MacDougall, 2010; McGillis Hall, Pedersen, Fairley, 2010; Tucker & Spear, 2006; Westbrook, Coiera, Dunsmuir, Brown, Kelk, Paoloni, & Tran, 2010a), in offices (e.g., Czerwinski, Horvitz, & Wilhite, 2004; González & Mark, 2004), and on aircraft flightdecks (e.g., Dismukes, Young & Sumwalt, 1998; Loukopoulos, Dismukes, & Barshi, 2001).

Research has shown that interruptions are disruptive to the completion of routine data-entry tasks in at least two ways. First, it takes time to recover from an interruption, in terms of the resumption lag needed to re-engage with the primary task (e.g., Cades, Boehm-Davis, Trafton, & Monk, 2011; Grundgeiger et al., 2010; Monk, Trafton, & Boehm-Davis, 2008; see Trafton & Monk, 2007, for a review). Second, interruptions increase the likelihood of errors being made on the primary task, in that important components of the task are either repeated or missed (e.g., LeGoullon, 2006; Li, Blandford, Cairns, & Young, 2008; Trafton,
Altmann, & Ratwani, 2011). In some settings, errors made in the execution of routine data-entry tasks can have disastrous consequences. For instance, interrupting nurses while they setup infusion pumps can increase the chances of medication delivery errors occurring (e.g., Trbovich et al., 2010; Westbrook, Woods, Rob, Dunsmuir, & Day, 2010b). In other settings, the consequences of errors are less severe and lead only to minor inconveniences and annoyance. For instance, an office worker might be interrupted by a telephone call while completing an expense form and forget to enter a relevant account code before submit it after finishing up with the call.

Given that the consequences of making an error can vary considerably between different workplace settings, we investigate whether people can adopt compensatory strategies to avoid making errors when they are costly and whether there are interventions that can be put in place to encourage such strategies to reduce the likelihood of error following an interruption. In doing this we consider whether there is a link between how quickly a task is resumed following an interruption (i.e., the duration of the post-interruption resumption lag) and the likelihood that an error is made. Despite an abundance of work investigating the effect of interruptions on routine task performance, there has been a surprising lack of discussion on the link between resumption lag and error. This is because studies that have focused on the time course of the recovery from an interruption have tended not to report separate resumption lag data for correct and incorrect resumptions (e.g., Altmann & Trafton, 2007; Hodgetts & Jones, 2006a, 2006b; Monk et al., 2008; Trafton et al., 2003). In contrast, a separate body of work has focused explicitly on errors made in the
execution of routine tasks, but these studies did not report any resumption lag data (e.g., Byrne & Bovair, 1997; Li et al., 2008; Trafton et al., 2011). Thus it is presently unclear whether task resumption errors are associated with longer or shorter resumption lags.

In the following section, we give a brief overview of Altmann and Trafton’s (2002) memory for goals theory. This theory is important as it provides a useful theoretical framework for thinking about how people recover from an interruption, allowing for predictions to be made about the likely relationship between resumption lag and error that we go onto test experimentally in this paper.

Memory for Goals

Altmann and Trafton’s (2002) memory for goals theory has emerged as an important framework for understanding how people resume a task following an interruption. The theory assumes that during goal-directed behavior people maintain a mental representation of their intentions to accomplish a task along with temporary information associated with that task (i.e., relevant problem state information, see Salvucci & Taatgen, 2011, for an extensive discussion). At the onset of an interruption, goal-relevant information must be temporarily stored in memory so that it can later be retrieved once the interruption has been dealt with. The theory assumes that the time taken to retrieve a suspended goal is dependent on its activation in memory. Based on activation-based accounts of human memory (e.g., Anderson, 2007; Anderson & Schooler, 1991), the theory also assumes that memory elements decay over time unless they are actively rehearsed. This time-based decay component of the theory leads to a number of
predictions that have been extensively tested in the literature, for instance, that longer interruptions will lead to longer resumption lags (e.g., Altmann & Trafton, 2002; Hodgetts & Jones, 2006a, 2006b; Monk et al., 2008).

While the duration of resumption lag is an important metric, resumption accuracy also matters. Unfortunately, the relative infrequency of errors has presented a barrier to studying human error experimentally (Gray, 2004; Sellen & Norman, 1992). Significant progress has been made in recent years developing experimental tasks that elicit high enough error-rates from participants to study human error in the laboratory (e.g., Back, Cheng, Dann, Curzon & Blandford, 2006; Byrne & Bovair, 1997; Byrne & Davis, 2006; Gray, 2000; Li et al., 2008). This has allowed for some of the predictions made by memory for goals about errors to be experimentally tested.

Memory for goals attributes resumption errors to noise in an activation-based memory retrieval process. The theory assumes that representations leftover from previously completed goals might, on occasion, become more active than the most recently suspended goal. Given that it is the goal representation with the highest activation that is retrieved, a previously completed goal might be retrieved instead of the correct goal. This leads to the task being resumed at the wrong point. From this perspective one might assume that resumption errors should be associated with longer resumption lags because the activation level of the retrieved item is closer to the interference level (i.e., it has a lower activation value in memory, see Altmann & Trafton, 2002, pp. 46). There is some empirical support for this idea: Monk et al. (2008) found that participants made more errors on a VCR programming task and
resumed more slowly following longer interruption periods. This finding appears to be broadly consistent with memory for goals theory, in that longer periods between rehearsals of a goal reduce the likelihood of successful retrieval.

Trafton et al. (2011) have focused on sequence errors made in the execution of routine data-entry task. Sequence errors occur whenever the steps in a specific task sequence are executed in the wrong order (Reason, 1984). For instance, people might repeat a previously executed step or skip over a step in a task sequence (Norman, 1981). In their experiment, Trafton et al. interrupted participants immediately after they had completed a subtask. This meant that participants had to keep track of where they were in the task sequence during the interruption period because they could not rely on the interface to provide reliable cues to help them to resume the task in the correct place. As expected, Trafton et al. found that interruptions led to an increase in the frequency of resumption errors being made compared to baseline (non-interruption trials). Interestingly though, most of the incorrect actions were in close proximity to the correct step (i.e., participants were most likely to either skip one step ahead or repeat the previously executed step). This suggests that participants in Trafton et al.’s study had a good idea of where they were in the task sequence. It is however unclear whether resumption errors were associated with longer or shorter resumption lags because Trafton et al. do not report resumption lag data.

Finally, memory for goals assumes that some of the negative effects of interruptions can be mitigated by active goal rehearsal. The theory assumes that rehearsal works to bolster the activation of a goal in memory making it more readily accessible at the point of retrieval. This means that tasks can be resumed
faster and there is a decreased likelihood of error. The theory does not make any strong commitments about when rehearsal must occur: It can be before an interrupting task is attended to (Trafton, Altmann, Brock, & Mintz, 2003) or it can occur concurrently with the interrupting task (Salvucci, Monk, & Trafton, 2009). It is however often difficult to know precisely if and when people are choosing to rehearse goal-relevant information. One solution to this issue is to elicit verbal protocols from participants so as to infer their strategy (as was the case in Trafton et al., 2003). An alternative approach that we use here is to use interrupting tasks that are cognitively demanding so as to inhibit the opportunity for rehearsal as much as possible (Cades, Boehm-Davis, Trafton, & Monk, 2007).

So far we have detailed a theoretical framework for describing how people resume after interruptions. We have also described a number of studies that have shown that interruptions take time to recover from, in the form of resumption lag. Longer resumption lags have usually been interpreted as being indicative of more disruption to the primary task, increasing the likelihood of error. However, despite extensive exploration of these measures of resumption performance, none of the studies outlined above reported separate resumption lag data for correct and incorrect task resumptions. We seek to address this important gap in the literature and report two experiments that investigate the relationship between resumption lag and error rate and the extent to which this relationship is mediated by strategic speed-accuracy tradeoffs.
Overview of the Experiments

In Experiment 1 we varied the time-cost penalty associated with making a resumption error following an interruption. In this way we hoped to investigate whether tasks were resumed more or less quickly depending on the relative cost of making an error. In addition, we explore whether errors are associated with faster or slower resumptions. In Experiment 2 we build on this finding by forcing slower task resumptions by way of a post-interruption lockout that stops the participant from resuming the primary task immediately after dealing with the interruption. In this way we were interested in finding out whether more or fewer errors would be made. Together these studies are designed to provide insights into the relationship between resumption lag and accuracy following an interruption.

In both of the experiments participants were interrupted while working on a routine sequential data-entry task, called the Doughnut task (Li et al., 2008). We chose this task for two reasons: (1) it has similarities to the practical routine data-entry tasks discussed in the introduction (e.g., programming an infusion pump to give a prescribed quantity of medication, completing an expense report using a spreadsheet application), and (2) it is a well-documented task that has been used extensively in previous research.

The task shows similarities with practical routine data-entry tasks; take the example of programming an infusion pump to give a prescribed quantity of medication to a patient. In this task, values are taken from a prescription sheet and programmed into the pump following a specific procedure using interface controls. This is mimicked in the Doughnut task by having participants read off a
series of to-be-entered value from a central display and enter it following a specific procedure using various interface controls.

A further benefit of the Doughnut task is that previous studies show that even well trained participants tend to make errors at a rate sufficient for analysis (e.g., Ament, Cox, Blandford, & Brumby, 2010; Hiltz, Back, & Blandford, 2010; Li et al., 2008; Ratwani, McCurry, & Trafton, 2008; Trafton et al., 2011), thus overcoming the issue of infrequent errors in experimental settings (Gray, 2004; Sellen & Norman, 1992). One of the reasons why the Doughnut task elicits errors is because of the intentional minimization of cues both during routine programming and after interruption. Previous studies have shown that cues both before interruptions (Altmann & Trafton, 2004; Trafton, Altmann, & Brock, 2005) and after interruptions (Chung & Byrne, 2008) have the potential to improve resumption performance, as such, minimizing them is an effective way of eliciting errors. One consequence of this decision is that there may be potential issues when generalizing from this task to other data-entry tasks that are performed on interfaces that do provide such cues. We return to this in detail in the General Discussion.

In terms of the interrupting task, it is well known from workplace studies that interruptions can lead to errors because they place additional demands on people and can take them away from their current activity, making it difficult to resume afterwards. For instance, a nurse might be summoned by an alarm while attending to a patient (Grundgeiger et al., 2010), while an office worker might have to answer the phone while working on a spreadsheet (González & Mark, 2004). To approximate this, participants in our study were interrupted partway
through working on the Doughnut task and asked to solve a series of mental arithmetic problems as quickly as possible. During this interruption period participants could not see or work on the Doughnut task. Following the interruption period, participants have to recall what it is they should do next. This would be difficult because the Doughnut task interface does not provide explicit cues to aid successful resumption and the high cognitive load of the interrupting task would limit active goal rehearsal during the interruption period itself (Cades et al., 2007). Hence, we could expect participants to make frequent resumption errors following an interruption.

**Experiment 1**

Across different settings the consequences of making an error can vary from a mere inconvenience through to a catastrophe. One might generally assume that people are sensitive to the cost of errors and will try to avoid making errors when they carry severe consequences. Yet surprisingly a number of previous studies have shown that people cannot be easily motivated to make fewer errors in the execution of routine procedural tasks (e.g., Back et al., 2006; Byrne & Davis, 2006; Chung & Byrne, 2008).

In Experiment 1 we investigate the effect that varying the cost of making an error has on post-interruption resumption lag and error rate. We explore this question in the context of our experiment by manipulating the time-cost penalty incurred when errors are made. In a low-cost condition errors incurred a very small time-cost penalty in that participants were simply notified of errors when they occurred but no further action was required. In contrast, errors made in a high-cost condition incurred a significant time-cost penalty, in that participants
were first prohibited from interacting with the system for a period of 20-seconds before they had to restart the primary task from the beginning of the trial. We assumed that this difference in time-cost penalty between conditions would influence participants' task resumption behavior, such that participants in the high-cost condition might adopt compensatory strategies to try and guard against making resumption errors.

One possible strategy participants might adopt to avoid making resumption errors in the high-cost condition is increased goal rehearsal. As outlined above memory for goals suggests that errors can be avoided through active goal rehearsal before or during the interruption period. Altmann and Trafton (2002) assume that when a goal is suspended it is strengthened through an active rehearsal process so that it can be easily retrieved at a later point in time. This change in rehearsal strategy would mean that the suspended goal would be retrieved more readily from memory, resulting in fewer errors and shorter resumption lags in the high-cost condition. While it would be very difficult to completely rule out this possibility, increased goal rehearsal seems unlikely in the present experiment for two reasons. First, there is no interruption warning to allow for effective pre-interruption rehearsal (Trafton et al., 2003), and second, the interrupting task itself is cognitively demanding so limiting the opportunity for active goal rehearsal during the interruption period (Cades et al., 2007).

An alternative way that participants might improve their task resumption accuracy is to give more time up to remembering where in the task sequence they were before resuming (i.e., a change in retrieval strategy). It is known that
retrieval accuracy improves with increased time (Dosher, 1976; Reed, 1973, 1976). Thus, we might expect that a rational response from the participant would be to allow more time before acting in the high-cost condition in order to avoid costly resumption errors. In this case we would expect to see longer resumption lags and fewer errors made in the high-cost condition. In contrast, when resumption errors are not so costly, a rational response from the participant might be to simply make a ‘fast guess’ (Ollman, 1966) and act on the first thing that comes to mind, since making an error carries little penalty. In this case we would expect to see shorter resumption lags and more errors made in the low-cost condition. These predictions are tested in the following experiment.

Method

Participants. Twenty-four participants (15 female) were recruited from the UCL Psychology Subject Pool, with a mean age of 21.6 years (SD = 3.4 years). Participants were randomly assigned to either the high- or the low-cost condition (12 participants in each condition). The experiment took approximately 70 minutes to complete and participants were paid £7 (approx. $11) for their time. None of these participants took part in Experiment 2.

Materials. The Doughnut and Packing tasks were implemented in Visual Basic 6. Software ran on a Dell Optiplex machine with 1 GB of RAM running Microsoft Windows XP. An optical mouse was used, set at the ‘medium’ speed via the system control panel. Materials were presented on a 17 inch TFT monitor set at a resolution of 1024 by 768 pixels. Participants were seated approximately 60cm from the monitor.
In the primary Doughnut task (adapted from Li et al., 2008), participants are required to process a specific order for doughnuts using a task interface. Figure 1 shows a screen shot of the interface for the Doughnut task. As can be seen in the figure, the interface is subdivided into a number of frames (e.g., Order Sheet, Dough Port, Puncher, Froster, Sprinkler, Fryer, Selector). We explain the role of each of these components in turn.

The central order sheet gives details on the specific type of doughnuts to be entered on a given trial. Two orders are given on each trial, where each is defined by a series of attribute-variable pairs (e.g., Quantity is 27, Dough is Crispy, Shape is Diamond, Frosting is Vanilla, Sprinkle is None). The order sheet remains present throughout the trial allowing the participant to check on it whenever they like.

To process a given order, the participant must enter values into the appropriate component of the interface. Each of these subtasks must be executed in following a specific sequence: (1) Dough Port, (2) Puncher, (3) Froster, (4) Sprinkler, and (5) Fryer. The final step in the task sequence is to click on the (6) Process button. This ends the trial and gives the participant feedback on what doughnuts were made.

To begin working on a given subtask, the participant must first activate it by clicking on the appropriate button in the Selector panel (which can be seen on the far right of the Doughnut task interface). After activating the subtask, the participant can enter the relevant information from the order sheet for that subtask in the appropriate interface frame. For example, working on the dough port, the participant would select the “crispy” radio-button and enter “27” in the
RECOVERING FROM AN INTERRUPTION

quantity textbox for the first order and then select the “original” radio-button and enter “49” in the quantity textbox for the second order. After this the participant had to mentally calculate and then enter the sum of the two quantities into the total textbox (i.e., 27 + 49 = 76). When all of the relevant information for a subtask is entered, the participant would click on the OK button. Pressing the OK button would cause all of the just-entered information to be erased from the interface, returning it to its initial (blank) state. This ensures that the interface does not provide any place-keeping cues to help the participant determine which subtasks has been completed and which has not; instead, the participant is required to remember where in the task sequence they are.

While working on the primary Doughnut task, participants were occasionally interrupted for a period of 30 seconds and required to complete a secondary mental arithmetic task, called the Packing task. The Packing task was designed to be cognitively demanding and required participants to calculate how many boxes of a given capacity would be needed to pack a quantity of doughnuts for delivery. For example, participants might be told that they needed to pack 14 doughnuts, and that they could use boxes that can hold either four or two doughnuts each. Participants had to give a solution to this problem (e.g., 3 x 4-box and 1 x 2-box could be used to pack 14 doughnuts). We used a 30-second interruption period based on the results of previous research (e.g., Monk et al., 2008), which has shown that this causes the kinds of resumption interference effects that we are interested in while still being practical in an experimental setting.
For each interruption trial there were two interruptions. Interruptions occurred immediately after a participant had completed a subtask, and there was no interruption warning. Interruptions could occur at one of the four possible interruption points in the primary task sequence, see Table 1.

Upon resuming the Doughnut task, the participant had to remember which subtask they had completed immediately prior to dealing with the interruption. A post-interruption resumption error was counted whenever the participant failed to click on the appropriate button in the Selector panel associated with next subtask to be executed in the task sequence. The interface did not provide any visual cues or state information to aid successful task resumption. The motivation for doing this was to increase the likelihood that participants would make errors, allowing for statistical comparisons to be made between conditions.

**Design.** A between-subject design was used to vary the cost of making sequence errors in the Doughnut task. A sequence error occurs whenever subtasks are completed in the wrong order. Whenever participants made a sequence error in the *high-cost* condition, an error message was shown on the monitor saying “Machine broken. Please wait for repair.” This error message remained on the screen for a period of 20-seconds during which time the mouse cursor changed to an hourglass and the Doughnut machine interface did not accept any user input. After this 20-second wait period was over participants were informed that they would have to re-enter all of the target values for the current trial again. In the *low-cost* condition, participants were notified of their error by an audible beep and the correct component was named on the order
sheet. This information was removed once the participant had pressed the correct button in the Selector panel. Critically, although errors in the low cost condition revealed the correct step to be executed next, errors nonetheless had to be corrected by the participant and so carried a time-cost relative to the situation where the participant did not make a resumption error in the first place.

Given that the focus of this study was on whether varying the cost of making a post-interruption resumption error would lead to a change in resumption lag, a between-subjects design was appropriate to avoid the risk of asymmetric transfer effects between conditions (e.g., Poulton, 1982). Put another way, a within-subjects design would have been inappropriate here because there is a risk that experiencing one condition first might lead to the adoption of a persistent resumption strategy that might be used in the other condition. Critically, this strategy might be different to that which might have been adopted had the participant never experienced the first condition. Hence we argue that a between-subjects manipulation is more appropriate as it avoids such transfer effects between conditions.

The main dependent variables of interest were the frequency of resumption errors following an interruption and the time taken to resume the primary Doughnut task. As outlined above, following an interruption the participant should click on the appropriate button in the Selector panel to select the next subtask to be executed. A post-interruption resumption error is any action other than this correct action. We only consider the first error made at any given resumption point rather than the total number of errors made. In
terms of resumption time, we consider the time taken to click on a button in the Selector panel once the Doughnut task interface appeared at the end of the interruption period.

**Procedure.** Participants were informed that they would be required to perform a routine sequential data-entry task while being occasionally interrupted by a secondary task, and that we were interested in how people recover from an interruption. After explaining the purpose of the study, informed consent was gathered from the participants in line with the procedure approved by the UCLIC Ethics Committee, which adheres to the Code of ethics and conduct of The British Psychological Society.

Participants were next given training on how to perform both the Doughnut and Packing tasks. To guarantee that participants knew how to perform the Doughnut task correctly, they completed at least three practice trials and did not progress to the experimental trials until they had completed at least two error-free trials.

In the main part of the study, participants completed 12 experimental trials and they were told that they could leave as soon as they had completed all of these trials correctly. Participants were informed that they would occasionally be interrupted while working on the Doughnut task, and that they would be required to work on the Packing task for a period of 30 seconds. Participants were instructed to solve as many Packing task problems as they could in the time available before resuming the Doughnut task. Interruptions occurred on six of the 12 experimental trials, and for each interruption trial there were two
separate interruptions (see Table 1 for details). The order of interruption trials was randomized.

Participants received different instructions on the consequences of making an error when programming the Doughnut task depending on which cost condition they were assigned to. Participants in the *high-cost* condition were told that errors on the Doughnut task would ‘break the machine’; that they would have to wait 20-seconds for it to be fixed; and that after this period they would have to restart the current trial from the beginning. Participants in the *low-cost* condition were told that they would be notified of any errors that were made on the Doughnut task and that these should be fixed. It should be noted that while in the *high-cost* condition, where trials were restarted whenever an error was made, the number of interruptions remained fixed regardless of the number of times that a particular trial was attempted. Specifically, if a participant made an error after the second interruption on a trial, then they would restart the trial, but there would be no further interruptions. If however a participant made a resumption error on the first scheduled interruption in a trial, then they would be interrupted only at the point of the second scheduled interruption.

The study took approximately 70 minutes to complete, and participants were given the opportunity to take a short break after completing half of the experimental trials. Participants were debriefed at the end of the study and payment was given.
Results

The main dependent variables of interest were the frequency of resumption errors and the time taken to resume the primary Doughnut task after dealing with an interruption. For statistical analysis an independent two-sample $t$-test was used, and effects were judged significant if they reached a .05 significance level.

To determine whether participants had learnt how to execute the task sequence for the Doughnut task given the training received, we first consider the percentage of sequence errors that were made on the non-interruption control trials. For each participant, there were a total of 36 opportunities for errors to occur (each participant completed six non-interruption control trials, each made up of six subtasks). Across the six non-interruption control trials, participants made very few sequence errors ($Mdn = 1$, range: $0 – 2$). There was no difference in the mean error-rate between conditions for the non-interruption trials ($M = 2.78\%$, $SD = 1.18\%$ vs. $M = 2.55\%$, $SD = 1.43\%$, for the high- and low-cost conditions, respectively), $t < 1$. This suggests that participants could execute the Doughnut task sequence correctly most of the time if they were left uninterrupted to work on it.

We now turn our attention to the error-rate and resumption lag data for the critical (interruption) trials. Each participant completed six interruption trials, each with two interruptions per trial. This meant that for each participant there were 12 interruption points. We consider the percentage of errors made at these resumption points and the average resumption lag in each condition. (Note: that the number of interruption points was the same for each condition. That is, even though in the high-cost condition a trial was restarted whenever an error was made, the
number of interruptions per trial did not change.) It was found that proportionally fewer errors were made by participants following an interruption in the *high-cost* condition (*M* = 10%, *SD* = 8%) than by participants in the *low-cost* condition (*M* = 24%, *SD* = 14%), *t*(22) = 2.91, *p* < .01, *d* = 1.19, 95% CI [4, 24]. In terms of mean resumption lag, it was found that participants were slower to resume the primary task following an interruption in the *high-cost* condition (*M* = 5.25s, *SD* = 0.23s) than participants in the *low-cost* condition (*M* = 3.87s, *SD* = 0.21s), *t*(22) = 15.40, *p* < .001, *d* = 6.29, 95% CI [1.19, 1.57]. Figure 2 shows the distribution of resumption times for both correct and incorrect resumptions for each condition. It can be seen in the figure that across both conditions resumption errors were associated with shorter resumption lags.

**Discussion**

Experiment 1 was designed to investigate how varying the cost of making an error affects task resumption behavior. The results show that when resumption errors carried a significant time-cost penalty, participants made fewer errors and resumed the primary task more slowly than when errors incurred a minor penalty. This finding suggests that participants were making speed-accuracy tradeoffs in deciding how quickly to resume the primary task based on the value of making an error. How people recover from an interruption has primarily been thought of in terms of a memory-based process in which the episodic context necessary to perform the task is retrieved from memory (Altmann & Trafton, 2002; Trafton & Monk, 2008). Given this memory-based framework, that participants made fewer errors with increasing resumption time is consistent
with the idea that memory retrieval accuracy increases as the time allowed for retrieval increases (e.g., Dosher, 1976; Reed, 1973, 1976).

An alternative explanation of these data is that participants in the high-cost condition might have changed their rehearsal strategy. Altmann and Trafton (2002) suggests that active rehearsal can bolster the activation of a suspended goal, making it more likely to be successfully retrieved from memory, and in less time. It seems unlikely that participants in the present study were able to rehearse prior to dealing with the interruption because there was no interruption warning. Another possibility is that participants might have rehearsed the suspended goal while performing the secondary Packing task (Salvucci et al., 2009). While it is very difficult to rule out this possibility without having gathered verbal protocols (as in Trafton et al., 2003) or data regarding performance on the interrupting task, it seems unlikely because the interrupting task was cognitively demanding and should have limited the opportunities for goal rehearsal (Cades et al., 2007). A further possibility that is more difficult to rule out is that participants could have rapidly encoded each action more deeply in the high-cost condition as they worked on it. However, all of these goal strengthening accounts would appear to be inconsistent with the present data because they predict that a more active goal should be retrieved more readily, leading to shorter resumption lags in the high-cost condition. This was not the case however. Instead the results suggest that participants were making speed-accuracy tradeoffs in terms of their retrieval strategy in order to minimize errors. We return to the issue of how a memory-based account might explain these data in more detail in the General Discussion.
That participants could be motivated to make fewer resumption errors is reassuring but also quite striking. This is because a number of previous studies have shown that people cannot be easily motivated to make fewer errors in the execution of routine procedural tasks (e.g., Back et al., 2006; Byrne & Davis, 2006; Chung & Byrne, 2008). However, these studies focused on whether people can be motivated to avoid making a special kind of procedural slip known as a postcompletion error (Byrne & Bovair, 1997). The postcompletion error occurs when people forget to execute a task step that comes after the main task goal has been achieved. The explanation for this error has focused on the idea that when the user’s primary goal has been achieved this acts as a kind of ‘false completion signal’ that leads the user to move on to their next activity rather than executing the post-completion step. In contrast, the resumption errors observed here are different. They occur because people incorrectly recalled what it is they were doing prior to the interruption, and so restarted the task in the wrong place. In this way, resumption errors reflect an error in memory retrieval, and our results suggest that this is subject to strategic speed-accuracy tradeoffs. One prediction that follows from this hypothesis is that enforcing longer task resumption periods should allow people more time to actively recall what it is they were working on before the interruption period and so reduce the likelihood of error. In the following study we test this hypothesis by evaluating the effectiveness of an enforced post-interruption lockout system that prohibits faster task resumptions.

**Experiment 2**

The results of Experiment 1 show that when resumption errors were costly, participants chose to resume the primary task more slowly and made fewer
resumption errors. In Experiment 2, we build on this finding by controlling how quickly participants can resume the primary task to see what effect this has on accuracy. We use the same general task setup as used in Experiment 1, with the exception that on half of the interruption trials a post-interruption lockout followed the end of the interruption period. During this system lockout period, participants could look at, but not interact with, the task interface. We expect participants to make fewer resumption errors when forced to resume the primary task more slowly. This is because the post-interruption lockout should prohibit the kind of fast, inaccurate task resumptions that were associated with errors in Experiment 1.

**Method**

**Participants.** Twenty-four participants (13 female) were recruited from the UCL Psychology Subject Pool, with a mean age of 23.3 years ($SD = 4.8$). The experiment took approximately 60 minutes to complete and participants were paid £7 (approx. $11) for their time. None of these participants took part in Experiment 1.

**Materials.** The experiment used the same task setup as the low-cost condition in Experiment 1, with the exception that on half of the interruption trials a post-interruption lockout system prohibited participants from resuming the primary task for a 10-second period. During this lockout period, participants viewed a surrogate interface of the Doughnut task (see Figure 3). This surrogate interface contained the subtask frames but had all interface controls removed. The controls were removed so that participants would clearly know that they were in the lockout period and that they could not interact with the interface. We
thought it important to show the general spatial layout of subtasks, rather than a blank screen, to give participants the opportunity to think about where they were in the task structure during the lockout period as this is known to aid task resumption (Ratwani & Trafton, 2008). Participants were not given any explicit instructions to do this, however. The end of the lockout period was signaled by the interface retuning to its default state (i.e., the input controls were made visible in each of the compartment frames of the Doughnut task interface). At this point participants were free to resume the primary task as normal.

**Design.** A within-subjects design was used to manipulate what happened following an interruption. In the *interruption-only* condition participants were interrupted while working on the primary Doughnut task and required to work on the Packing task for 30 seconds. In the *interruption+lockout* condition participants were prohibited from resuming the primary Doughnut task for a period of 10-seconds following an interruption. In addition, there was a series of non-interruption control trials in which participants completed a doughnut order without interruption. These provided a measure of baseline performance on the primary Doughnut task. As before, the dependent variables were the frequency of post-interruption resumption errors made in each condition and the time taken to resume the Doughnut task.

**Procedure.** The procedure was the same as in Experiment 1 expect participants did at least two practice trials (where at least one had to be error-free) before moving on to the main experimental trials. After training, participants completed eight interruption trials (four *interruption-only* trials and four *interruption+lockout* trials), along with four non-interruption control trials.
The order of trials was randomized. For each of the interruption trials there were two separate interruptions, the timing of which was randomly selected from the list of possible interruption points (see Table 1). Participants were told that on some of the interruption trials they would be 'locked-out' from the Doughnut task for a 10-second period following the interruption. Participants were told that they could use this time to look at, but not interact with, the surrogate interface. However, participants were not given any explicit instructions as to what they should do during the lockout period. The experiment took approximately 60 minutes to complete and participants were given the opportunity to take a short break after completing half of the experimental trials.

**Results**

We first consider the frequency of sequence errors that occurred in the non-interruption control trials to determine whether the participants could execute the task sequence for the Doughnut task given the training received. For each participant, there were 24 opportunities in total for errors to occur (each participant completed four non-interruption trials, each made up of six subtasks). Across the four non-interruption control trials, participants made very few sequence errors ($Mdn = 1$, range: $0 - 3$). The mean error-rate was $5.9\%$ ($SD = 3.5\%$). This suggests that participants were able to execute the task sequence correctly most of the time given the training received.

We next consider data from the interruption trials to determine how the enforced 10-second lockout period affected performance. Figure 4 shows the distribution of resumption times for both correct and incorrect resumptions for
both the interruption+lockout and interruption-only trials. Results show that proportionally fewer resumption errors were made in the interruption+lockout condition \((M = 8\%, \ SD = 10\%)\) than in the interruption-only condition \((M = 22\%, \ SD = 15\%)\), \(t(23) = 5.33, \ p < .001, \ d = 1.08, \ 95\% \ CI [9, 20]\). For the analysis of the resumption lag we consider the effective resumption lag between conditions. That is, to correct for the fact that, in the lockout trials, participants could not resume the task until after the lockout period had passed we subtracted 10 seconds from the resumption lags from that condition, leaving the resumption lags from the control condition unaltered. Participants had a shorter effective resumption lag in the lockout+interruption condition \((M = 2.54s, \ SD = 0.25s)\) than in the interruption-only condition \((M = 3.81s, \ SD = 0.71s)\), \(t(23) = 9.04, \ p < .001, \ d = 2.39, \ 95\% \ CI [.98, 1.56]\).

**Discussion**

The results of Experiment 2 show that a post-interruption lockout period can effectively reduce the likelihood of resumption errors being made following an interruption. The lockout period prohibited participants from resuming the primary task for 10 seconds after dealing with a 30-second interruption. Because fewer resumption errors were made, participants likely used this enforced lockout period to recall where they were in the task sequence prior to dealing with the interrupting task. Once the lockout period was over, participants resumed the task relatively quickly, suggesting that they had been preparing to execute a planned action (i.e., they could have prepared a mouse movement over the relevant interface control). However, unlike in the interruption-only condition, these fast resumptions led to fewer errors being made.
The results of the current experiment are broadly consistent with a number of previous studies that have also shown benefits to using lockout periods to encourage users to use more memory-intensive strategies (Gray, Sims, Fu, & Schoelles, 2006) and plan extended sequences of actions (O’Hara & Payne, 1998, 1999; Svendsen, 1991). Of particular relevance here is the work of Morgan, Patrick, Waldron, King, and Patrick (2009), who found that increasing the access costs for looking at to-be-entered information (i.e., similar to our order sheet) during a routine data-entry task facilitated resumption behavior after an interruption. This is because when it was difficult for participants to access information more of it was encoded into memory meaning that it did not have to be re-checked after dealing with an interruption. The results of the current study further support the idea that encouraging users to adopt a memory-intensive strategy, for instance, by taking the time to plan their actions before executing them, can facilitate resumption behavior and reduce the likelihood of error. Future experiments might manipulate the duration of the lockout period in order to parameterize the boundary conditions of this effect. We return to this in more detail in the General Discussion.

The current experiment also helps address a lingering question on the role that rehearsal might play in facilitating task resumption behavior in this task setup. As was discussed above, we believe that there was limited opportunity for rehearsal before or during the interruption period. Because of the design of Experiment 2 we can be quite confident that the observed decrease in error-rate in the lockout condition was not due to a change in rehearsal strategy. This is because following an interruption participants were unaware as to whether they
were going to be free to resume the task immediately or whether they would be locked out of the system for a further 10-seconds. That the lockout condition led to a significant reduction in error-rate could only have been due to what occurred during the post-interruption lockout period.

One concern with the current experiment is that because a within-subjects design was used participants did fewer non-interruption control trials in Experiment 2 than in Experiment 1 (four vs. six trials, respectively) and that this might have affected how well participants performed the Doughnut task. There is some evidence of a marginal increase in error-rate on non-interruption trials between the two experiments (5.9% vs. 2.7%, for E2 and E1, respectively). On closer inspection though this difference most likely reflects a consequence of how error-rate is calculated. Across the two experiments there was not a difference in the total number of errors made; participants tended to make just one error across all of the control trials (i.e., the median error-count was 1). This suggests that reducing the number of trials per condition did not have a detrimental impact on how well participants could perform the Doughnut task in this case.

**General Discussion**

Interruptions are common across a variety of workplace setting in which people have to perform data-entry tasks and increase the likelihood that errors are made. However, the consequence of making an error can vary considerably from one setting to another. The results of Experiment 1 show that people are sensitive to the cost of making a resumption error and adapt their behavior accordingly. In particular, we found that when errors incurred a significant time-
cost penalty, participants choose to resume the primary task more slowly, and as a result made fewer resumption errors. Longer resumption lags have usually been interpreted as being indicative of more disruption to the primary task which, given the lower error rate, appears not to be the case here. This work therefore raises the question of whether longer resumption lags following an interruption may be beneficial in terms of reducing the occurrence of error.

Experiment 2 built on the idea that longer resumption lags reduce the likelihood of error by implementing a 10-second system lockout period that encouraged participants to reflect on how they were going to resume the task. It was found that this lockout period led to a significant reduction in resumption errors compared to when participants were free to resume the primary task at their own (faster) pace. This result suggests that forcing people to slowly re-engage with a task following an interruption reduces the likelihood that resumption errors are made. However, further work is obviously needed to assess the practical value of post-interruption lockouts in operational environments outside the laboratory setting used here. One concern would be how to design a system in such a way that the lockout cannot easily be worked around. For instance, an expected workaround might be for a user to press a button to ‘unlock’ a device while turning their attention to another task until the lockout period has elapsed. In this scenario the value of the lockout might be missed since attention is not on the main task during the lockout period. Another important concern is that in some situations it is critical that tasks are completed quickly, as is the case in many healthcare settings. Some consideration would therefore need to given to whether the benefit given by a lockout mechanism in
terms of reduction in errors might be mitigated by the effect of slowing users down in the completion of their work. One way forward here would be for future research to investigate the effectiveness of lockouts of different durations. The lockout used in our study was quite long at 10-seconds. For instance, it may be the case that far briefer lockouts are also effective in some settings.

Taken together, the results of these two experiments clearly demonstrate that longer resumption lags following an interruption reduce the likelihood that resumption errors are made. This result that longer resumption lags are beneficial to task performance is actually somewhat surprising when considered against the context of the dominant theoretical account of how people recover from interruptions. As outlined above, memory for goals assumes that the episodic context necessary to resume an interrupted task must be retrieved from memory (Altmann & Trafton, 2002, 2007; Trafton & Monk, 2008; Trafton et al., 2011). We contend that Altmann and Trafton’s (2002) memory for goals theory, as currently implemented, does not easily offer an account for why longer resumption lags lead to fewer resumption errors being made. This is because the ACT-R memory retrieval mechanism (Anderson, 2007), which Altmann and Trafton (2002) build on, assumes that memory representations decay over time and that as time passes these memory representations become more difficult to retrieve, both in terms of retrieval accuracy and latency.

There are a number of possible avenues for extending the memory for goals theory so that it might explain the results reported here. One approach would be to alter the memory retrieval mechanism so that increasing the time allowed for retrieval would result in fewer resumption errors being made. One
way that this might be done is that, with increased time, multiple episodic memory representations might be slowly recovered in order to resume the task correctly. This idea is broadly consistent with accumulator models (e.g. Smith & Vickers, 1988) and random walk models (e.g. Stone, 1960; Link, 1975; Ratcliff, 1978), which generally assume that the more time that is given over to a decision process the more likely the outcome is to be correct. Indeed, there have been proposals for just such accumulator retrieval mechanisms to be integrated within ACT-R (Van Maanen & Van Rijn, 2007, see also Anderson, 2007). Detailed computational modeling work, which is beyond the scope of this present article, is required to assess the viability of such a change to the memory for goals retrieval mechanism.

A different approach would be to extend memory for goals theory beyond mere-memory retrieval functions and incorporate the perceptual re-encoding of the interface that might occur during the resumption period. Trafton et al. (2011) discuss the idea that when resuming a task people might choose between a fast-but-fallible strategy of retrieving the relevant problem-state information from memory and a slow-but-accurate resumption strategy of perceptually re-encoding information from the interface (see also Gray et al., 2006; Gray & Fu, 2004). In this way, Trafton et al. (2005, 2011) argue that taking the time to re-encode such information from the environment can support accurate task resumption. To lend support to this idea, Salvucci (2010) used an ACT-R model to explain resumption lag data from an interruption study reported by Iqbal and Bailey (2005). Salvucci argues that the long resumption lags observed by Iqbal and Bailey could not be explained by memory-retrieval processes alone but had
to reflect slower perceptual re-encoding processes. As discussed above, the interface for the Doughnut task does not provide explicit cues to aid task resumption so as to elicit errors. However, it is plausible to assume that participants might have attempted to mentally re-construct where they were in the task sequence by stepping through the interface prior to resuming. Indeed, this was the intention of providing the surrogate interface during the lockout period in Experiment 2. Such cognitive re-construction would take time but may explain the increased accuracy seen here. The value of resuming a task slowly then is that it gives people time to re-engage with the task, for instance, by re-encoding information from the task interface, so as to avoid making an error.

Finally we give some consideration to whether some of the decisions that were made when designing our experiments might limit the generalizability of our findings to different settings and tasks. First, the Doughnut task did not provide explicit cues to aid task resumption. Some caution is therefore required when generalizing from the results reported here to data-entry tasks performed on interfaces that provide explicit cues to help the user keep track of what values have been entered and what state the system is in. Removing these interface cues as we did in our experiments likely lead to a substantial increase in error-rate because participants were forced to use their memory to keep track of their progress in the task. That said, it does seem reasonable to assume that even when an interface provides explicit cues to support task resumption, our general claim that longer task resumptions reduce the likelihood of error would hold because people would still have to perceptually re-encode this information from the interface to make use of it and this would take time (see, Gray et al., 2006;
Gray & Fu, 2004; Trafton et al., 2011). Further work is necessary to explore this issue in detail.

Second, participants in our study had a lack of control on how and when they choose to deal with the interrupting Packing task. In particular, interruptions were of a fixed duration and participants had no choice but to suspend the primary task immediately at start work on the interrupting task. This approach is common across a large body of experimental studies that have investigated interruptions in laboratory settings (e.g., Ament et al., 2010; Cades et al., 2011; Hiltz, Back, & Blandford, 2010; Li et al., 2008; Monk, Trafton, & Boehm-Davis, 2008; Ratwani, McCurry, & Trafton, 2008; Trafton et al., 2003; Trafton, Altmann, & Ratwani, 2011). While there are clearly situations in which interrupting tasks are suddenly thrust upon people, there are also many others in which people have far more control over how and when they choose to deal with an interrupting task. For instance, a ringing cell phone can be ignored and the call returned later. Studies have been conducted that give participants such discretionary control in deciding when to deal with an interrupting task. For instance, Salvucci and Bogunovich (2010) have shown that when given the choice people will defer an interrupting task until they reach a natural breakpoint in the primary task, such as the completion of a subtask (see also Bogunovich & Salvucci, 2011). But as we have shown here dealing with an interruption at a subtask boundary leaves open the possibility that a resumption error is made when the primary task is resumed.

Third, participants in our studies had no choice but to try and resume the primary task where they left off following an interruption. In reality though
many data-entry tasks can often be restarted. For instance, observational studies conducted in hospitals have revealed that nurses will often choose to reboot an infusion pump rather than change the values that have already been partially programmed into it (Furniss, Blandford, & Mayer, 2011). This decision to restart the task rather than recover it partway through seems appropriate given the potentially fatal consequences associated with infusion pump programming errors (Institute for Safe Medication Practices Canada, 2007; Thimbleby & Cairns, 2010). Future work might explore in a more controlled experimental setting whether people would be more likely to restart routine data-entry tasks when the cost of making an error is increased. This would point to yet another kind of speed-accuracy tradeoff that might occur in resumption strategy and would extend our initial finding that taking more time to resume a task following an interruption mitigates the risk of making an error.

In conclusion, interruptions during the execution of routine data-entry tasks are disruptive and can increase the risk of error. The results of this paper suggest that the likelihood of resumption errors can be reduced by encouraging/forcing users not to resume their primary task quickly.
References


RECOVERING FROM AN INTERRUPTION


Proceedings of the 25th BCS Conference on Human-Computer Interaction
(pp. 197-204). Swinton, UK: British Computer Society.


Table 1

*Steps in the Doughnut task and possible interruption positions*

<table>
<thead>
<tr>
<th>Procedure</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Select Dough Port, Enter Parameters, Click OK</td>
</tr>
<tr>
<td>2. Select Puncher, Enter Parameters, Click OK</td>
</tr>
<tr>
<td><strong>Possible Interruption Point</strong></td>
</tr>
<tr>
<td>3. Select Froster, Enter Parameters, Click OK</td>
</tr>
<tr>
<td><strong>Possible Interruption Point</strong></td>
</tr>
<tr>
<td>4. Select Sprinkler, Enter Parameters, Click OK</td>
</tr>
<tr>
<td><strong>Possible Interruption Point</strong></td>
</tr>
<tr>
<td>5. Select Fryer, Enter Parameters, Click OK</td>
</tr>
<tr>
<td><strong>Possible Interruption Point</strong></td>
</tr>
<tr>
<td>6. Click the Process button</td>
</tr>
</tbody>
</table>
**Figure 1.** The Doughnut task interface (adopted from Li et al., 2008).

Numbered boxes represent the sequence in which data should be entered to each compartment from the central order sheet. Before a compartment can be used it must be selected by pressing the relevant button in the Selector panel (highlighted box). Sequence errors were defined as any action in which the participant failed to select the relevant button in the Selector panel at the appropriate point in the task sequence.
Incorrect resumptions
Correct resumptions

Figure 2. Distribution of resumption lags for correct and incorrect task resumptions in the (a) low-cost and (b) high-cost condition from Experiment 1.
Figure 3. The surrogate interface of the Doughnut task that participants saw during the 10-second lock-out period in Experiment 2. The surrogate interface removed all of the interface controls so that participants know that they could not interact with the interface. The subtask frames remained in place to give participants the opportunity to think about where they were in the task structure during the lockout period.
Figure 4. Distribution of resumption lags for correct and incorrect task resumptions in the (a) interruption-only and (b) interruption+lockout condition from Experiment 2.