Investigating the Effects of Using Checklists on Routine Task Performance under Different Working Memory Load Conditions

Aaron Chan

University College London

Department of Psychology

Course: Research Project (3901)

Student’s Contact Address: aaron.chan@ucl.ac.uk

Word Count: 8167 (Excluding References)

Due Date: February 25, 2011
Abstract

Interruptions have deleterious effects on people’s performance on routine procedural tasks, leading to higher error rates. According to the activation-based goal memory (AGM) model, errors occur due to competition between memory representations. Research has shown that appropriate interventions can reduce the occurrence of such errors. An experiment, involving 56 participants and of a mixed design type, was carried out to determine the benefits of checklist use under different working memory loads. It was hypothesised that (i) a checklist would be beneficial for task resumption following an interruption, (ii) higher working memory load would be detrimental to task performance, and (iii) a greater improvement in performance would be observed when using a checklist under high working memory load. Results do not support the hypotheses. Various experimental limitations that could explain the results, and practical implications of the current research, are discussed.
Interruptions are commonplace in both office (Czerwinski, Horvitz, & Wilhite, 2004) and healthcare settings (Chisholm, Collinson, Nelson, & Cordell, 2000), which may lead to errors in task execution. Such errors can have disastrous consequences, particularly in situations that are safety critical in nature. It is worth noting that these errors can occur even after training, and that people termed experts in carrying out any procedure in question are not immune to them (Back, Brumby, & Cox, 2010). Moreover, it has been found that enhanced training on the procedure may not reduce the occurrence of such errors (Back et al., 2010). These findings highlight the notion that as operators of machines and devices, humans are susceptible to influences that may hinder the correct execution of a task. Human error has thus been termed an inevitable consequence of the inclusion of the human operator in the procedural loop (Hales & Pronovost, 2006). Nonetheless, although human error may be impossible to eliminate completely, researchers have strived to identify various means by which human error can be mitigated, mainly focusing on the reduction of error rates following an interruption during task execution. This paper provides an overview of research on the impact of interruptions on routine procedural tasks, as well as to build on the findings derived from human error research by investigating the usefulness of external aids such as checklists during task execution.

Several findings have been borne out of research on routine procedural errors. For example, it has been found that increased working memory load leads to higher error rates (Ament, Cox, Blandford, & Brumby, 2010). Moreover, specific error types have also been identified. In procedures that
require subtasks to be completed in a specified order, interruptions increase the likelihood that these subtasks are executed in the wrong order, and such errors have been termed sequence errors (Reason, 1984). Various types of sequence errors exist, depending on the nature of the actions executed following an interruption (Trafton, Altmann, & Ratwani, 2009). Perseveration errors occur when a previous step is repeated, whereas anticipation errors occur when the next correct step to be executed is skipped altogether. Some procedures also require a postcompletion step to be executed at the end of the task sequence, and it has been found that people are more likely to omit this step if they are interrupted just prior to its execution (Li, Blandford, Cairns, & Young, 2008). These observations demonstrate that human error, while evident in natural and applied settings, may be studied usefully in more controlled environments, helping to provide a deeper understanding on how people respond to and deal with interruptions during task execution.

Various models have been formulated to explain the aforementioned observations, in the context of routine procedural action. The interactive activation network (IAN) model postulates that various schemas compete for activation (Cooper & Shallice, 2006). Such activation is derived from contextual or environmental triggers, or from other related schemas. However, any particular schema will have to be above a specific threshold in order to be activated. Additionally, the IAN model predicts that errors in procedural tasks occur due to distractions and insufficient allocation of attentional resources (Cooper & Shallice, 2000). Essentially, any noise introduced can result in the disordering of schemas, as well as the incorrect selection of the appropriate schema when multiple schemas are present. Such variability attributable to
noise can lead to sequence errors. On the other hand, the simple recurrent network (SRN) model postulates that environmental features are responsible for the activation of a set of input units (Botvinick & Plaut, 2006). This activation is propagated along to a set of hidden units where activation is recirculated, and then on to a set of output units that determine the action to be performed. Various connection weights are used to encode sequential attractors which represent the task sequence. Introducing noise may then result in the network drifting to an incorrect task sequence, or sequential attractor, that resembles the next step, leading to a sequence error. Hence, while the IAN model emphasises the role of attentional resources in sequence errors, the SRN model stresses the changing internal state of the system that leads to information losses (Botvinick & Bylsma, 2005). These models illustrate that noise, possibly in the form of distractions from external sources, interfere with information processing, leading to suboptimal task performance.

Despite the apparent success of the IAN and SRN in explaining sequence errors, empirical work on routine procedural tasks has identified the activation-based goal memory (AGM) model (Altmann & Trafton, 2002) as a more useful framework for understanding the effects of interruptions in the execution of goal-related tasks. Essentially, the AGM model postulates that when goal-directed behaviour is concerned, people maintain an internal representation of the intention to achieve the goal, as well as other temporary information that is necessary for the successful completion of the task. During an interruption, the cognitive system must attempt to store all information related to the task for subsequent retrieval during task resumption following the interruption. In line with activation-based accounts of human memory
(Anderson et al., 2004), the AGM model states that the amount of time required to retrieve the stored information is contingent on its level of activation in memory. Moreover, this memory representation is subject to decay with time unless active rehearsal is performed. The AGM model thus predicts longer resumption times following extended interruptions. Additionally, the introduction of noise may permit older memory representations of completed goals to reach activation levels that are higher than that of the suspended goal. This may then lead to the incorrect retrieval of the older memory instead of the currently suspended goal, resulting in a resumption error. However, this can be avoided if the person actively rehearses the memory representation of the suspended goal during the interruption, which may also translate to shorter resumption times following the interruption.

Several studies involving problem-solving tasks (Altmann & Trafton, 2002; Hodgetts & Jones, 2006b) and routine sequential tasks (Altmann & Trafton, 2007) have produced findings that largely agree with the AGM model. In one study (Hodgett & Jones, 2006b), participants were asked to solve the Tower of London task, while occasionally being interrupted by a secondary task. The duration of this secondary task was manipulated. It was found that participants took longer to resume the Tower of London task following the interruption when the duration of the secondary task was increased. In a separate study (Monk, Trafton, & Boehm-Davis, 2008), participants were trained in a routine sequential task involving the programming of a VCR. Once again, the participants were occasionally interrupted by an unrelated secondary task. It was found that resumption times varied with the duration of
the secondary task, with longer interruption times being more detrimental to task performance. These experiments have largely supported the predictions offered by the AGM model with regards to the impact of interruptions on procedural tasks.

Recent studies have, however, utilised newer paradigms in order to elicit high enough error rates for useful investigation and analyses of human error (Li et al., 2008). One example is the Wicket Doughnut task (Li et al., 2008), which is essentially a routine procedural task requiring participants to process orders of virtual doughnuts. The orders can only be processed if the subtasks are completed in a certain order. Participants may also be interrupted by an unrelated secondary mental arithmetic task during the course of processing the doughnut orders. Li et al. (2008) found that interruptions increased the likelihood that participants resumed the task at the wrong point. These are essentially the sequence errors described earlier (Reason, 1984), where a step in the sequence is repeated or left out (Norman, 1981). Trafton and colleagues (2009) modified the original version of the Wicket Doughnut task by removing all available information from the task interface following the successful completion of each subtask. In line with previous studies, it was found that interruptions led to higher error rates. Additionally, Trafton et al., (2009) attempted to explain the difference between perseveration and anticipation errors, both of which are considered sequence errors. Perseveration errors occur when older memories are incorrectly selected due to transiently higher activation levels. In contrast, anticipation errors arise when a planned step is assumed to be completed, resulting in task resumption ahead of the correct step. Trafton et al. (2009) also
discovered that participants were more likely to commit errors in close proximity to the correct step. These findings are largely in agreement with the AGM model mentioned above, that errors arise due to the competition between competing memory representations during retrieval. Moreover, due to the gradual decay of memory traces with time, the probability of error should be highest at sequences closest to the correct step. This prediction was observed in the aforementioned study (Trafton et al., 2009).

It can thus be seen that there has been a heavy emphasis on memory processes in explaining task resumption following interruptions in routine procedural tasks. Indeed, Gawande (2010) has stressed that human memory and attention are not impervious to interruptions and distractions, which serve to impede task performance. Researchers have thus sought to identify memory-based solutions to minimise the occurrence of human error. Back and colleagues (2010) investigated the effects of an enforced lockout period following an interruption in a routine procedural task. Using the Wicket Doughnut task (Li et al., 2008), it was found that participants were less likely to commit sequence errors following the enforced lockout period. Results showed that lockout periods reduced the incidence of sequence errors by up to 64%. Essentially, during the enforced lockout period, participants were only allowed to look at but not interact with the task interface prior to resuming the primary task. It was suggested that participants were able to use this time to mentally retrace their steps prior to the interruption, as well as to plan ahead in preparation for task resumption (Back et al., 2010). Indeed, O’Hara and Payne (1999) also found that on a task involving processing a word document, participants completed the task more efficiently following a lockout.
This was on the premise that participants used the time during the lockout period to anticipate future actions required to complete the task in the most efficient manner. The findings from the aforementioned studies fit nicely with the predictions of the AGM model (Altmann & Trafcon, 2002). With more time at their disposal, participants could have made attempts to strengthen memory representations of the suspended goal, shielding it from interference at the point of retrieval due to higher activation levels as a result of the strengthening process. This decreases the likelihood of retrieving the incorrect memory representation, hence leading to a reduction in error rates. Therefore, it can be seen that interventions aimed at improving memory retrieval following task interruptions may be beneficial at reducing the incidence of sequence errors on routine procedural tasks.

While the use of enforced lockout periods can help reduce error rates following an interruption, a more accurate retrieval of the memory trace required for the correct execution of subsequent steps in a task may be achieved through the use of a checklist. A checklist is essentially a list of action items arranged in a sequential order, which affords the user to monitor task progress while the steps are being completed (Hales & Pronovost, 2006). Checklists can thus be considered a memory retrieval aid, providing guidance as to which steps must be completed in order to accomplish the task, while simultaneously serving as a verification tool allowing the user to identify which steps remain outstanding in the task sequence (Hales & Pronovost, 2006). Existing literature is laden with research findings that suggest the importance of checklist usage at the workplace. Given that pilots and astronauts are exposed to inherently high-risk work environments, much of the research on
the use of checklists have focused on the fields of aviation and aeronautics (Hales & Pronovost, 2006). Aviators consider the checklist as a form of flight protocol, which must be strictly adhered to. Failure to do so constitutes a protocol violation, or pilot error (Helmreich, 2000). Indeed, studies have shown that 33% of jetliner accidents are caused by pilots deviating from established operational procedures (Sears, 1965). Checklists exist for various phases of a typical flight, including preflight checks, engine start checks, take off checks, as well as landing and shutdown checks (United States Air Force (USAF), 2006). These ensure consistency and coherence in normal flight operations. Checklists have also been employed in emergency situations, allowing the flight crew to systematically address the problem without compromising safety (Boorman, 2001). Given the importance of checklists in ensuring flight safety, it is no wonder why the aviation industry adopts a serious view on the use of checklists. Despite the success of checklist usage in error reduction, aircraft manufacturers have strived to constantly seek improvements to existing checklist systems. For example, implementation of the Boeing 777 Electronic Checklist further decreased errors by up to 46% when compared to previous paper-based versions of the checklist (Boorman, 2001). In 2008, a British Airways Boeing 777 crashed upon landing at London’s Heathrow Airport (Department of Transport, 2008). Investigators determined that ice crystals were responsible for clogging fuel lines during landing, leading to a sudden loss of thrust. The Federal Aviation Administration and Boeing jointly developed a revised checklist to include supplementary procedures preventing the freezing of trace amounts of water in the fuel system on flights plying the polar routes (Federal Aviation
This checklist helped prevent another accident involving a Delta Air Lines Boeing 777 flight, which suffered an uncommanded engine rollback due to ice crystal formation in the fuel system (National Transportation Safety Board (NTSB), 2009). Hence, it can be seen that the use of checklists in the aviation industry has led to improved work processes and best practice adherence, which have largely improved flight safety standards (Hales & Pronovost, 2006).

Similar findings have also surfaced in the medical and healthcare industry. Wolff, Taylor, and McCabe (2004) found that using checklists in clinical care resulted in heightened compliance in various practices in response to patients admitted for acute myocardial infarction, leading to an improvement of key outcomes by as much as 55%. In critical care, using checklists to specify the withdrawal-of-life-support process led to fewer patients being administered with inappropriate resuscitation measures or medications prior to death (Hall, Rocker, & Murray, 2004). The use of checklists has also been found to reduce the infection rates through the administration of proper antibiotics, as well as increased operation success rates through improved standardisation of practices and better communication between staff (Gawande, 2010). Checklists have also been found to be effective in situations of exceptionally high stress, such as trauma response and management (Harrahill & Bartkus, 1990). Indeed, it has been mentioned that the use of a mental checklist by the attending trauma surgeon helped US congresswoman Gabrielle Giffords survive a gunshot wound to the head (Associated Press (AP), 2011). As with the aviation industry, checklists are
poised to exert greater influence on work processes amounting to more efficient practices in healthcare.

It is important to consider exactly how checklists serve as cognitive aids in procedural tasks. It has been suggested that the category superiority effect (Sharps, Wilson-Leff, & Price, 1995) may explain the usefulness of checklists. Essentially, the grouping of information in an organised manner has been found to improve recall, a phenomenon commonly referred to as chunking. Indeed, cockpit checks are broken down according to the different phases of flight, making it more manageable for the flight crew to handle all necessary procedures without compromising safety. Additionally, studies have shown that instructions presented as a list are often better understood and remembered when compared to continuous prose (Morrow, Leirer, Andrassy, Hier, & Menard, 1998). Thus, a checklist serves as a learning aid and a memory retrieval cue, allowing the user to execute the procedure in a more optimal and efficient manner.

Unfortunately, research on checklist use has been largely restricted to applied settings, and there is a dearth of such research in the laboratory. Investigating checklist use in more controlled settings should allow a more accurate description of the cognitive mechanisms behind the advantages of checklist use in routine procedural tasks. Given the potential of lockout intervention in reducing error rates (Back et al., 2010), it would be prudent to determine whether additional memory aids, such as checklists, can be used in conjunction with an enforced lockout to further reduce error rates following an interruption. An experiment, of a mixed design type, was carried out to investigate the benefits of using a checklist by extending the findings of Back.
and colleagues (2010) with the inclusion of a checklist during the lockout period following an interruption, under different working memory load conditions. The working memory load manipulation serves to ascertain the relative effectiveness of checklist use under potentially stressful working conditions. It was hypothesised that (i) a checklist would be beneficial for task resumption following an interruption, (ii) higher working memory load would be detrimental to task performance, and (iii) a greater improvement in performance would be observed when using a checklist under high working memory load. The between-subject independent variables were checklist use and working memory load, while the within-subject independent variable was the type of interruption schedule. The dependent variables were error rate and resumption lag. Results from the current study can be used to further the current understanding of how checklists help reduce error rates in routine procedural tasks, as well as how to assimilate checklist use in applied settings in order to improve workplace efficiency and best practice adherence.

Method

Participants

A total of 56 participants (30 male and 26 female) were involved in the experiment. The age range was 19 to 24 years, with a mean of 21.34 years ($SD = 1.40$).

Materials

The current study employed an adapted version of the Wicket Doughnut task (Li et al., 2008). Figure 1 illustrates the screenshot of the
Doughnut task interface. For each trial, participants were required to process an order of a certain quantity and type of doughnuts. The information necessary to complete the order was presented in an Order Sheet which was displayed in the centre of the interface for the duration of the trial. To successfully process an order, participants had to input the required information in the various compartments of the virtual doughnut machine. These subtasks had to be executed in a pre-defined sequence: (1) Dough Port, (2) Puncher, (3) Froster, (4) Sprinkler, (5) Fryer, and (6) Process. Clicking the “Process” button signaled the end of the trial.

In order to commence entering information for a subtask, participants had to initialise the relevant compartment by clicking the relevant button on the Selector panel, which was displayed on the right side of the task interface (see Figure 1). Following that, participants could then proceed on to entering the relevant information as detailed on the Order Sheet for that particular subtask. After entering the information, participants then clicked the “OK” button to complete the subtask. This also caused the recently entered information to be removed from the interface. Participants were thus required to keep track of their progress on the Doughnut task since the interface did not provide them with cues to determine exactly which subtasks had been completed.
On selected trials, participants working on the primary Doughnut task were interrupted by a secondary mental arithmetic task for 30 seconds. This was called the Packing task, as participants had to calculate the number of boxes of differing capacities required to pack a specified quantity of doughnuts. For example, participants had to decide that in order to pack 43 doughnuts using boxes with capacities of 4 and 9 doughnuts, they would need a total of 4 and 3 of such boxes respectively. The secondary Packing task was only inserted in between subtasks. Table 1 illustrates the possible interruption points on interruption trials. Note that the Packing task was randomly inserted into any 2 of the 4 possible interruption points. After the Packing task, participants had to resume the primary Doughnut task without any cues from the interface, but instead had to rely on memory in order to determine which was the next subtask to be executed. Additionally, on selected interruption trials, participants were presented with an enforced lockout period of 10 seconds during which they were presented with a surrogate Doughnut task interface with controls removed, and were unable to
resume the primary task. With regards to the working memory load manipulation, participants had to keep track of differing attributes of doughnuts being sold while they were processing the orders. This information was presented to participants at the bottom of the task interface.

Table 1

*Task description with possible interruption points indicated*

<table>
<thead>
<tr>
<th>Procedure</th>
<th>Possible Interruption Point</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Select Dough Port, Enter Parameters, Click OK</td>
<td></td>
</tr>
<tr>
<td>2. Select Puncher, Enter Parameters, Click OK</td>
<td></td>
</tr>
<tr>
<td>3. Select Froster, Enter Parameters, Click OK</td>
<td><strong>Possible Interruption Point</strong></td>
</tr>
<tr>
<td>4. Select Sprinkler, Enter Parameters, Click OK</td>
<td><strong>Possible Interruption Point</strong></td>
</tr>
<tr>
<td>5. Select Fryer, Enter Parameters, Click OK</td>
<td></td>
</tr>
<tr>
<td>6. Click the Process button</td>
<td></td>
</tr>
</tbody>
</table>

The checklist used in this study appeared as a list of actions required to successfully process an order on the primary Doughnut task (see Appendix). The checklist was formulated following design guidelines, such as listing and grouping principles, as well as ensuring a logical flow of checklist items (Degani & Wiener, 1993). For example, the various subtasks were listed according to the pre-defined sequence which was required of the primary Doughnut task. Additionally, the relevant information that was to be entered for each component were grouped under the corresponding subtask heading,
allowing for easy access. The checklist was printed on an A4 size paper, within easy reach of participants requiring the use of the checklist.

**Design**

The current study was a mixed-design experiment that aimed to investigate the effects of checklist use on error rate and resumption lag under different working memory conditions. The between-subject independent variables were checklist use (present or absent) and working memory load (high or low), while the within-subject independent variable was the type of interruption schedule (no interruption, interruption only, interruption with lockout). The dependent variables were error rate and resumption lag. Participants were thus randomly assigned to one of the following 4 groups: “Checklist-High Load”, “No Checklist-High Load”, “Checklist-Low Load”, and “No Checklist-Low Load”. However, all participants were each presented with 12 trials (4 from each interruption schedule), and the order of presentation of trials was completely randomised.

**Procedure**

Participants were first provided with an overview of the current study, and were informed that they were required to perform a routine procedural task involving processing orders of virtual doughnuts. Additionally, participants were briefed that they might be occasionally interrupted by a secondary mental arithmetic task during the execution of the primary task, and that the study was primarily interested in how people respond to interruptions. Prior to the commencement of the Doughnut task, participants completed a web-based backward digit recall task (www.cognitivefun.net/test/11), which served
as a working memory capacity test. Essentially, after being presented with a random string of digits, participants had to enter the numbers in reverse order. Starting with only 3 digits, participants could only progress to the next level after 2 correct attempts per digit span. Using a ladder procedure, the working memory span for each participant was calculated. This procedure aimed to identify any potential outliers with respect to working memory capacity.

Subsequently, participants were presented with at least 3 practice trials, which served as a form of training on the Doughnut and Packing tasks. To ensure a minimum level of competency on the tasks, participants were required to complete at least 2 error-free practice trials before proceeding on to the experimental trials.

With regards to the experimental trials, participants were informed that they would be presented with 12 trials that were similar in nature to the practice trials, and that they might be interrupted by the Packing task on selected trials for a period of 30 seconds. Participants were instructed to solve as many Packing task problems during this period. Additionally, participants were informed that following the Packing task, they would be presented with an enforced lockout period of 10 seconds on selected interruption trials, during which they would not be able to continue with the primary Doughnut task. In all, 8 out of 12 experimental trials were interruption trials, of which half contained the 10 seconds enforced lockout period, and the order of presentation of all 12 trials was randomised.

Participants in the different experimental conditions were provided with specific instructions. Those in the “Checklist” condition were asked to refer to the checklist during the lockout period before resuming the primary Doughnut task.
task. Participants in the “No Checklist” condition were not equipped with this checklist. With respect to the working memory load manipulation, participants were informed that they had to continuously keep track of certain attributes of virtual doughnuts sold while they were processing the orders. In the “Low Load” condition, participants had to monitor the number of Original doughnuts being sold. In the “High Load” condition, participants had to monitor the number of Crispy and Round doughnuts being sold. Hence, it was presumed that participants in the latter condition were subjected to a higher working memory load, as they had to monitor more attributes compared to participants in the former condition. This manipulation was adapted from the study by Ament and colleagues (2010). The entire experiment took approximately 60 minutes to complete. Participants were allowed to take a 30 seconds break after completing half of the experimental trials. A thorough debrief was conducted at the end of the experiment, providing participants with an opportunity to clarify any doubts that they might have had.

Results

The dependent variables measured in the current study were error rate and resumption lag. An action was registered as correct when the participant selected the correct compartment on the Selector Panel. All other actions were defined as errors. The total number of registered errors was converted to a percentage, or error rate, based on the total number of possible opportunities for an error to occur. For each of the 3 types of trials (no interruption, interruption only, interruption with lockout), there were 24 opportunities for an error to occur. The resumption lag was defined as the time taken to recommence the primary Doughnut task after the end of the
secondary Packing task. In the case of trials with the enforced lockout period, the resumption lag was similarly defined as the time taken to resume the primary Doughnut task after the end of the lockout period. In both these cases, the time taken was measured between the end of the interruption or lockout period, and the clicking of the appropriate compartment on the Selector Panel.

**Working Memory Capacity**

With regards to working memory capacity, it was found that participants were matched on their performance on the backward digit recall task, $t(54) = 0.19$, $p > .05$, indicating that working memory capacity was similar across participants.

**Error Rate**

Analyses of error rates revealed that sequence errors were more likely to take place following an interruption, regardless of whether checklists were used, in both high and low working memory load conditions. A 3 (interruption schedule) X 2 (checklist) X 2 (working memory load) mixed ANOVA was carried out on the error rates, with interruption schedule as the within-subjects measure, and checklist and working memory load as between-subjects measures. It was found that the main effect of interruption schedule was significant, $F(2, 104) = 7.41$, $p < .002$. Further analyses revealed that error rates for interruption only ($M = 18.90\%, SD = 15.81$) and interruption with lockout ($M = 20.24\%, SD = 14.16$) trials were significantly higher than no interruption trials ($M = 14.36\%, SD = 16.12$), but were not significantly different from each other. Indeed, comparing interruption and no interruption
trials alone, it was found that interruption trials ($M = 19.57\%$, $SD = 13.95$) had significantly higher error rates than no interruption trials ($M = 14.36\%$, $SD = 16.12$), $t(55) = 3.50$, $p < .001$. In other words, participants were more likely to make sequence errors following an interruption (see Figure 2), but the lockout period did not lead to a reduction in the error rate. There were no significant main effects for either the use of checklist or the working memory load manipulation, $F_s < 1$. All other interactions were also found to be not significant.

![Figure 2](image.png)

*Figure 2.* Mean error rates for trials with or without interruptions. Bars show standard error of means.

**Resumption Lag**

Analyses of resumption lag data revealed that participants resumed the primary Doughnut task faster when the interruption was followed by the enforced lockout period, regardless of checklist use and working memory load manipulations. A 2 (interruption schedule) X 2 (checklist) X 2 (working
A mixed ANOVA was carried out on resumption lag data, with interruption schedule as the within-subjects measure, and checklist and working memory load as between-subjects measures. It should be noted that only interruption trials were considered here as resumption lags were only measured following the secondary Packing task. There was a main effect of interruption schedule, $F(1, 52) = 19.90, \ p < .001$. Participants took less time to resume the primary Doughnut task following interruptions on trials with the enforced lockout period ($M = 5053.69 \text{ ms, } SD = 1820.18$) compared to trials without the enforced lockout period ($M = 6373.82 \text{ ms, } SD = 1935.98$) (see Figure 3). There was no main effect of the use of checklist, $F < 1$. There was also no main effect of working load manipulation, $F(1, 52) = 1.29, \ p > .05$. All other interactions were also found to be not significant.

![Figure 3](image.png)

*Figure 3.* Mean resumption lag times for interruption trials with and without the enforced lockout period. Bars show standard error of means.
Discussion

The current study investigated the usefulness of a checklist following task interruption under different working memory load conditions. It was hypothesised that (i) a checklist would be beneficial for task resumption following an interruption, (ii) higher working memory load would be detrimental to task performance, and (iii) a greater improvement in performance would be observed when using a checklist under high working memory load. The results do not support the hypotheses. It was found that participants were equally likely to make errors regardless of whether a checklist was used or not. Participants also showed similar performance levels in both the low and high working memory load conditions. Additionally, given that no interactions between checklist use and working memory load were observed, there was no greater improvement in performance when participants under high working memory load used a checklist.

Despite the lack of positive results with respect to the specific hypotheses, the current study has replicated findings from previous research on the effects of interruptions on routine procedural task performance. For example, the results show that sequence error rates were significantly higher following an interruption. According to the AGM model (Altmann & Trafton, 2002), such errors occur because the cognitive system struggles to store information crucial to task completion in memory for subsequent retrieval. As this information is subject to decay with time, the relevant representations in memory will be subject to interference from older memory representations of previously completed subtasks. The noise introduced increases the likelihood of the incorrect step being executed at the point of task resumption, leading to...
a perseveration error. Alternatively, the correct subtask may be skipped, as a result of the planned step being incorrectly remembered as having being already completed, leading to an anticipation error. The current study has established findings that are largely in agreement with previous research with regards to the detrimental effects that interruptions have on routine procedural action.

Additionally, the findings from the current study have revealed that participants took less time to resume the primary Doughnut task if an enforced lockout period was inserted immediately after the secondary Packing task. A study by Brumby, Cox, and Back (submitted) also generated similar findings, where participants had a shorter resumption lag when presented with an enforced lockout period immediately after the secondary interruption task. Essentially, the lockout period prevented participants from resuming the primary Doughnut task immediately after the secondary Packing task, allowing for a more accurate recall of the exact point in the task sequence in which they were interrupted (Brumby et al., submitted). Following the lockout period, participants could then resume the task quickly, following the additional preparatory steps taken during the lockout period. The AGM model (Altmann & Trafton, 2002) can be used to explain this observation. Although the relevant memory representations were subjected to decay and interference during the secondary Packing task, participants could use the enforced lockout period to actively rehearse the steps involved in the primary Doughnut task, allowing them to identify the correct subtask to be executed upon primary task resumption. This additional rehearsal would have increased the activation level of appropriate memory representations, ensuring that the
primary task was resumed at the correct point in the task sequence. The AGM model (Altmann & Trafton, 2002) thus provides a framework to explain the usefulness of enforced lockout periods in reducing the time required for task resumptions following interruptions in routine procedures.

However, findings from the current study show that despite the shorter resumption times observed on interruption trials with the enforced lockout period, there was no difference in sequence error rates between interruption trials with and without the enforced lockout. In other words, while participants were faster to resume the primary Doughnut task following the lockout, this did not lead to greater accuracy in task resumption. This observation may be reconciled with findings obtained from a study by Brumby and colleagues (submitted), which showed that participants were faster and more accurate in resuming the primary task following a lockout period. Brumby and colleagues (submitted) explained their findings using a speed-accuracy tradeoff framework, emphasising that lower error rates are usually associated with longer resumption times. According to this view, there should have been evidence in the current study showing that faster resumptions led to higher error rates. This trend was not observed, as the error rates remained the same despite shorter resumption times. However, it is important to note that the shorter resumption times were evident only in trials with the enforced lockout period following the interruption. This indicated that participants were using the lockout period strategically in preparation for primary task resumption. Hence, participants were able to resume the primary task expeditiously without sacrificing accuracy. This observation reiterates the
usefulness of lockout periods in mitigating the deleterious effects of interruptions on procedural task performance.

Results from the current study also showed that the use of a checklist did not assist in task resumption following an interruption, as evidenced by the similar error rates and resumption lag times regardless of checklist use. Several reasons could have accounted for this observation. The first concerns the conditions under which checklists may prove to be useful. Research has shown that checklists are particularly beneficial under situations with some level of complexity involved (Hales & Pronovost, 2006). Some examples include flight operations, critical medical care, and product manufacturing. People are more likely to be subjected to an increase in stress and fatigue under such circumstances, leading to a higher probability of errors made. It is possible that the Wicket Doughnut task (Li et al., 2008) may not be representative of such complex scenarios that warrant the use of a checklist. Another reason that could explain the ineffectiveness of checklist use in the current study concerns the mode of procedural learning involved. Checklists are usually employed to complement the learning of a particular procedure. For example, pilots are critically evaluated during flying training to ensure that they are competent in executing the various checks required to ensure safe flight operations. This usually involves the memorisation of such procedures with the assistance of a checklist. Subsequently, the checklist then serves as a memory aid to ensure that pilots adhere strictly to the established procedures. In the current study, participants were not required to memorise the checklist prior to the commencement of the experiment. Instead, the checklist was only presented to participants during the lockout periods just
prior to resuming the primary Doughnut task. It was assumed that the checklist could serve as an episodic memory retrieval cue to aid in task resumption. However, studies have shown that spatial memory can be used to facilitate task resumption (Ratwani & Trafton, 2008). Indeed, during the lockout period, the surrogate Doughnut task interface that was presented to the participants still retained the layout of the various compartments, despite having the controls removed. Nonetheless, the use of spatial memory to determine where in the task sequence an interruption occurred could have masked the effect of checklist use as an alternative episodic memory retrieval cue to aid in task resumption. A third reason that could explain the lack of usefulness of the checklist in the current study is related to the repetitiveness of the task. Given that participants were presented with a total of 12 experimental trials, practice effects could have led participants to be overly familiar with the task sequence under repeated exposure to the task. As a result, any effect that could explain the usefulness of the checklist would have been diminished as a result of task familiarity. The aforementioned accounts could explain why the checklist served no useful purpose in the current study.

Another intriguing finding was that the working memory load manipulation did not affect task performance. This was despite previous demonstrations that high working memory load leads to increased error rates during the execution of routine procedural tasks. For example, using the Wicket Doughnut task (Li et al., 2008), Ament and colleagues (2010) showed that high working memory load had a deleterious effect on task performance. Byrne and Bovair (1997) explained that high working memory load results in a decrease in the relative activation levels of the memory representations.
relevant to the task, and errors occur as a result of the heightened interference levels. However, Ament and colleagues (2010) found that although high working memory load increased overall error rates, device-specific steps were more negatively affected compared to task-specific steps. Device-specific steps are defined as actions necessary for the correct operation of the device, but may not contribute directly to the main goal. In contrast, task-specific steps are required irrespective of the device used, since they contribute directly towards the main goal. The current study made no attempt to differentiate between these two types of steps, and could thus explain why the working memory load manipulation was not observed to have an effect on task performance. The task could also have been too tedious for participants, given that both the secondary Packing task and the requirement to monitor the attributes of doughnuts sold required exceptionally high levels of information processing. It should be noted that the deleterious effect of high working memory load on procedural task performance observed in the study by Ament and colleagues (2010) was observed under considerably lower levels of task intensity, where participants only had to keep track of doughnuts bearing certain attributes on top of the primary Doughnut task. In contrast, participants in the current study had to grapple with the secondary interrupting Packing task that involved mental arithmetic problem solving. The task could therefore have been too demanding for an effect of working memory load to surface. The inclusion of a no working memory load control condition would have allowed for a more thorough analysis in determining whether the low working memory load condition itself was too difficult to begin with. In view of
these experimental limitations, it was thus possible that the detrimental effect of high working memory load was not observed in the current study.

The following limitations should also be noted. In this modified version of the original Wicket Doughnut task (Li et al., 2008), cues were removed from the task interface following each subtask. As such, participants had to rely solely on memory to monitor task progress, particularly after interruptions where errors are more likely to occur. This could be taken as a design flaw, since the interface should be used to provide cues indicative of task progress (Brumby et al., submitted). However, it should be noted that certain real life procedures involve devices that may not provide any cues, similar to what was employed in the current study. For example, a typical flight checklist may include items which involve the toggling of switches, or simply a visual check of an onboard aircraft system, such as fuel quantity. The former checklist step will provide visual feedback cues as the pilot will be able to confirm that the switches were indeed operated. In contrast, the latter step only requires that the pilot carry out a visual check of the system’s status without having to carry out any physical action or data input. Hence, this particular step does not involve the provision of any visual feedback cues to the user. The design employed in the current study could thus be an example of such a procedure. Participants were also not allowed to restart the entire procedure when faced with difficulties, and this could be considered as another experimental limitation (Brumby et al., submitted). Indeed, people often do have the option to restart a procedure in real life, particularly if precision and accuracy are of paramount importance. Hence, the current experimental design may not be representative of such routine procedures. Finally, apart from sequence
errors, other potentially important errors were not considered in this study. Errors such as those involving data entry may be particularly important in certain procedures. For example, Thimbleby and Cairns (2010) have stressed that number entry errors may have potentially lethal consequences in the medical field, such as during the use of an infusion pump in the administration of drugs. Such errors, however, were not considered in this study.

The aforementioned limitations could have accounted for the pattern of data observed in the current investigation. For example, the nature of the task may not have been suitable to investigate the effectiveness of using a checklist to improve task performance, as checklists have been found to be useful in more complex procedures (Hales & Pronovost, 2006). Moreover, the task could have been cognitively overly demanding to detect any effect of the working memory load manipulation employed here. This is an important dissociation to note, as task complexity does not equate to task intensity. In the current context, the Wicket Doughnut task itself can be considered a relatively simple procedure, but the additional manipulations required exceptionally high levels of concentration from the participants. Future research should account for this in the investigation of routine procedures. For example, checklists could be employed in experiments involving the use of more lengthy and complex procedures. Prior exposure to the checklist through learning sessions should also be explored, increasing the ecological validity of the experiment, as people are usually equipped with the checklist as a learning aid while training to carry out a procedure. More careful manipulations of working memory load should also be considered. For example, studies have shown that concurrent and sequential multitasking are
two qualitatively different processes (Salvucci & Taatgen, 2010). On one hand, concurrent multitasking involves multiple tasks that progress simultaneously with each other. In contrast, sequential multitasking involves a switching between tasks, where only one task is being attended to at any one time. The current study involved both concurrent and sequential multitasking, as participants had to perform the primary Doughnut task, interspersed with the secondary Packing task, while at the same time having to keep track of sales of doughnuts bearing certain attributes. However, as mentioned above, no effect of working memory load manipulation was observed. Ament and colleagues (2010) had found the effect of working memory load using only a concurrent multitasking design. Hence, experimental findings may be sensitive to the type of multitasking employed in the design of the task, and this should be taken into account in subsequent studies. Alternative paradigms could also be explored in the study of routine procedural action. For example, Ratwani, McCurry, and Trafton (2008) have found evidence that eye movements may be used to predict the incidence of error rates. Such measures may be coupled with existing experimental designs to provide more accurate estimates of error rates.

In summary, findings from the current study have important practical implications that can contribute to the understanding of how interruptions affect procedural task performance, and the additional safeguards people may take to mitigate human error. The results largely reaffirm the deleterious effects that interruptions have on routine procedures. This should lead to a greater emphasis on the use of various interventions to help minimise the incidence of human error in applied settings. For example, enforced lockout.
periods following interruptions during task execution may encourage people to retrace previously completed steps, as well as to prepare for the subsequent steps upon task resumption. Given the heavy emphasis on memory processes involved in the study of routine procedures (Altmann & Trafton, 2002), memory aids such as checklists could be employed in various work environments. Although the benefits of checklist use were not demonstrated in the current study, much research has shown the immense potential that checklist use has on improving work practices and ensuring safety, particularly in high intensity settings such as flying and critical medical care (Hales & Pronovost, 2006). It should also be noted that a checklist will only be as useful as the amount of effort put into its design process. By adopting a theory-driven approach to the study of how checklists may serve as an effective memory aid in highly stressful situations, the checklist design and implementation process should continually evolve, leading to the development of more effective measures that assist in task execution. For example, research has shown that the implementation of electronic checklists onboard the Boeing 777 led to a decrease in error rates when compared to conventional paper-based checklists (Boorman, 2001). More automated human monitoring and checklist procedures could be developed, coupling technology designed to predict errors and memory aids to assist in procedural task execution. This is in light of research that eye movements may be used to predict the incidence of errors (Ratwani et al., 2008). Such technology may be useful in certain work environments, such as the flight deck, where error prediction may be important in ensuring safety. Moreover, it has been found that checklist use may be considered as a system, involving complex
A. Chan

interactions between the operator and machine (Degani & Wiener, 1993). Strengthening this tightly coupled system should improve work efficiency while simultaneously reducing the likelihood of errors. However, it should be cautioned that excessive checklist use may have its drawbacks. For example, the excessive amount of checklists at the workplace may lead to checklist fatigue (Hales & Pronovost, 2006). In such a situation, the checklist impedes work performance, leading to a reduction in overall efficiency. Additionally, strict adherence to checklists may remove the need for professional judgment, especially when complex decisions need to be made. A drastic example would involve the use of an automatic-sensed checklist, which serves more like a built-in test function. It was found that pilots were less likely to detect anomalies in the system when such a checklist was used, hence highlighting the dangers of the over-reliance on checklists (Palmer & Degani, 1991). Nonetheless, given the immense potential of checklist use in applied settings, future research should focus on scenarios where the employment of checklists may be most productive, by reducing the incidence of human error. Indeed, Gawande (2010) has noted that the usefulness of checklists often goes unnoticed, and that this seemingly simple tool can bring about vast improvements in people’s lives.

Acknowledgements

The author would like to extend his gratitude to Dr. Anna Cox for her supervision and guidance, Dr. Jonathan Back for his advice in the formulation of the project idea, and Mr. Sandy Gould for his technical assistance in the programming of the experimental software.
References


A. Chan


A. Chan


Appendix

Wicket Doughnut Machine Checklist

Doughnut Machine Checklist

1. Receive next order
   - Choose dough type
   - Enter weight
   - Enter total weight

2. Select Dough Port
   - Choose dough type

3. Select Puncher
   - Choose hole shape
   - Select quantity
   - Specify dough type

4. Select Froster
   - Choose frosting
   - Slide rule (quantity)
   - Specify hole shape

5. Select Sprinkler
   - Choose topping
   - Select quantity
   - Specify topping

6. Select Fryer
   - Choose dough type
   - Enter amount of oil

7. Process order