Journal of Experimental Psychology: Applied
2008, Vol. 14, No. 4, 299–313

1076-898X/08/$12.00 DOI: 10.1037/a0014402

The Effect of Interruption Duration and Demand on Resuming Suspended Goals

Christopher A. Monk
George Mason University

J. Gregory Trafton
Naval Research Laboratory

Deborah A. Boehm-Davis
George Mason University

The time to resume task goals after an interruption varied depending on the duration and cognitive demand of interruptions, as predicted by the memory for goals model (Altmann & Trafton, 2002). Three experiments using an interleaved tasks interruption paradigm showed that longer and more demanding interruptions led to longer resumption times in a hierarchical, interactive task. The resumption time profile for durations up to 1 min supported the role of decay in defining resumption costs, and the interaction between duration and demand supported the importance of goal rehearsal in mitigating decay. These findings supported the memory for goals model, and had practical implications for context where tasks are frequently interleaved such as office settings, driving, emergency rooms, and aircraft cockpits.

Keywords: interruption, goals, interleaved tasks, memory

For most people, dealing with interruptions is not a problem to be overcome as much as it is an inevitable part of life. In fact, the ability to "multitask" is considered a desirable job skill by many employers, which is not surprising given that, on average, workers shift between tasks every 3 min (Gonzalez & Mark, 2004). By shifting between tasks every few minutes, it appears that people are managing interruptions by interleaving them with their primary tasks. For example, many people engage in conversations through instant message applications while working on other projects on the computer. The need to understand how interruptions and multitasking behaviors impact performance in the workplace has spawned several studies in recent years (e.g., Czerwinski, Horvitz, & Wilhite, 2004; Iqbal & Horvitz, 2007; McFarlane & Latorella, 2002).

A study that investigated the use of instant message communications in the workplace found that conversations lasted nearly 4.5 min on average, with exchanges every 15 s or so (Isaacs, Walendowski, Whittaker, Schiano, & Kamm, 2002). The study also showed that workers who heavily used instant messaging covered multiple topics in an exchange, and frequently shifted attention to other tasks while conversing. Avrahami and Hudson (2006) found that 92% of messages were responded to within 5 min, with 50% responded to within 15 s. They also found that students and interns used instant messaging at about double the rate of researchers in the sample.

The impact of interruptions is not merely an inconvenience for people going about their work and home lives. Interruptions can have devastating consequences. Multiple-plane crashes have been attributed in part to interruptions to the preflight checklists pilots perform prior to take-off (National Transportation Safety Board [NTSB], 1969, 1988). Studies have also shown that interruptions can affect driving safety (Monk, Boehm-Davis, & Trafton, 2004) and emergency room care (Chisholm, Collison, Nelson, & Cordell, 2000, Chisholm, Dornfeld, Nelson, & Cordell, 2001). Given the prevalence of interruptions and their potential for harmful consequences, it is not surprising that researchers have turned their attention to understanding how people perform when interrupted.

Although interruptions research dates back to the 1920s when Zeigarnik (1927) reported that people recalled details of interrupted tasks better than uninterrupted tasks, there was a long gap in experimental studies of interruptions until those conducted by Kreifeldt and McCarthy (1981) and Gillie and Broadbent (1989). The Kreifeldt and McCarthy and Gillie and Broadbent studies concluded that people performed postinterruption tasks more slowly compared to preinterruption performance. They also found that people made more errors in postinterruption performance; results corroborated initially by Cellier and Eyrolle (1992) and later by Zijlstra, Roe, Leonova, and Krediet (1999).

Subsequent interruptions studies primarily focused on determining the characteristics that make interruptions disruptive (see McFarlane & Latorella, 2002, for a comprehensive review). Several characteristics have been shown to affect primary task performance, including task similarity to the primary task (Cellier &

Christopher A. Monk and Deborah A. Boehm-Davis, Department of Psychology, George Mason University; J. Gregory Trafton, Naval Research Laboratory, Washington, DC.

This work was partially supported by the U.S. Office of Naval Research to J. Gregory Trafton: N0001405WX2001 and N0001405WX30020. The views and conclusions contained in this document should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U.S. Navy.

We thank Yi-Fang Tsai and David Shin for their assistance in data collection, and the Arch Lab members for their many helpful comments on earlier drafts.

Correspondence concerning this article should be addressed to Christopher A. Monk, Department of Psychology, George Mason University, 4400 University Drive, MSN 3F5, Fairfax, VA 22030-4444. E-mail: cmonk@gmu.edu

Eyrolle, 1992; Czerwinski, Chrisman & Rudisill, 1991; Edwards & Gronlund, 1998; Oulasvirta & Saariluoma, 2004), interruption complexity (Cades, Trafton, Boehm-Davis, & Monk, 2007; Gillie & Broadbent, 1989; Hodgetts & Jones, 2006b; Zijlstra et al., 1999), the relatedness of the primary and interruption tasks (Cutrell, Czerwinski, & Horvitz, 2001; Zijlstra et al., 1999), control over interruption onset (McFarlane, 2002), and the availability of primary task retrieval cues (Cutrell et al., 2001; Czerwinski, Cutrell & Horvitz, 2000). Unfortunately, some of the findings in these studies have been contradictory. For example, some found that interruptions slowed down performance on the primary task (Gillie & Broadbent, 1989), and some found that performance was faster when interrupted (Zijlstra et al., 1999). Speier, Valacich, and Vessey (1999) found that decision making on simple tasks was aided by interruptions but hindered for complex tasks. Recent evidence suggests that primary task performance is not the only victim of interruptions; secondary task performance can suffer in addition to primary task performance (Einstein, McDaniel, Williford, Pagan, & Dismukes, 2003; McFarlane, 2002).

These studies are useful for understanding how people might better deal with interruptions to work more efficiently, but they lack a cohesive theoretical approach to understand how people manage the multiple and temporary goals that result from interruptions. Altmann and Trafton (2002) introduced such a theoretical model for memory for goals that is particularly suited for the study of interruptions. This model has been tested in multiple-interruptions studies (Altmann & Trafton, 2007; Hodgetts & Jones, 2006a, 2006b; Li et al., 2006; Monk et al., 2004; Trafton, Altmann, Brock, & Mintz, 2003), and it is the basis for the predictions for interruption recovery in this study.

Memory for Goals

Altmann and Trafton's (2002) memory for goals model is a formal model of goal encoding and retrieval in memory. In their work, Altmann and Trafton successfully applied this model to simulating reaction time and error data from the Tower of Hanoi, a task that depends heavily on suspension and resumption of goals during problem solving. The suspension and resumption of goals is a fundamental aspect of interrupted task performance. For example, a person's current "train of thought" (primary task goal) when writing a report must be halted or suspended when an instant message arrives from an important source. As with all conversations, there is turn taking in instant message conversations that allows the person to return attention to the report-writing task while waiting for a response. With each shift of focus, the person must regain the suspended train of thought to resume writing the report. Because the model was developed to handle such suspended and resumed goals, it is well suited to predict the impact of interruptions on primary task resumption.

The memory for goals model is based on the activation model of memory items and is instantiated within the ACT–R cognitive architecture (Anderson, 1993; Anderson et al., 2004; Anderson & Lebiere, 1998). The fundamental processing assumption in this theory is that when central cognition queries memory, the chunk that is most active at that instant is returned. Returning to the example above, the writer's current goal or action is that with the highest level of activation at that moment in time. It is this goal that directs behavior (Anderson & Lebiere, 1998; Newell, 1990). Altmann and Trafton (2002) used an adapted version of ACT–R's Base Level Learning Equation (Anderson & Lebiere, 1998) to determine levels for goal memories. Within the ACT–R framework, a memory element's base-level activation represents its activation without any associations or cues (Anderson & Lebiere, 1998; Lovett, Reder, & Lebiere, 1999). A goal's retrieval history plays a significant role in its activation level, and therefore when it directs behavior (see Altmann & Trafton, 2002, for a detailed explanation). Frequently sampled goals will have higher levels of activation, as will recently encoded or retrieved goals. For example, the report writer in the above example will have more success in resuming a suspended train of thought if it was the focus of attention just before the interruption (*recency*) or for long periods before the interruption (*frequency*). The activation time course for a goal is depicted in Figure 1.

The effect of interruptions on task performance can be examined with the memory for goals framework (Altmann & Trafton, 2002) as a theoretical explanation of the determinants of goal activation and therefore, behavior. For example, when a goal is interrupted by another goal the original goal memory will immediately begin to suffer activation decay (assuming that the interrupting task engages the cognitive resources that would otherwise be used to rehearse such information). The time required to resume the suspended goal after the interruption is directly related to its level of activation (Altmann & Trafton, 2002). Goals that have been suspended for longer periods will have decayed to lower activation levels and therefore will take longer to resume, assuming no intervening rehearsal. In other words, the report writer will have greater difficulty resuming the suspended train thought when the instant message conversation persists for longer periods without opportunity to shift focus back to the report. Therefore, the memory for goals model predicts that longer interruptions should result in longer times to resume the primary task (or goal). Hodgetts and Jones (2006b) recently demonstrated support for this prediction.

Interruption Duration

Interruption duration has produced mixed results in the literature. Earlier studies that manipulated duration failed to show an

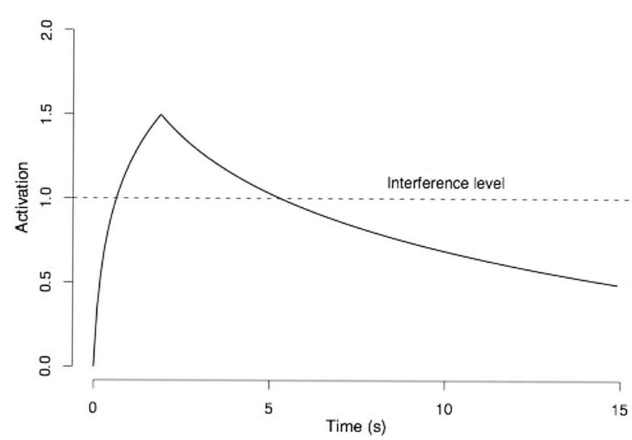

Figure 1. The time course of activation of a new goal (solid line) and the interference level from old goals (dashed line). Adapted from Altmann and Trafton (2002).

effect (Einstein et al., 2003; Gillie & Broadbent, 1989; Li et al., 2006). Recently, Hodgetts and Jones (2006b) were successful in finding an interruption duration effect using a Tower of London task when testing predictions from Altmann and Trafton's (2002) memory for goals model. Participants were prompted after three moves to click on a "mood" button appearing at the bottom of the computer display to open a mood checklist task (interruption). The interruptions lasted either 6 or 18 s, consisting of one or three nonrepeated mood checklists. In addition, the "mood" button indicated the length of interruption by noting the number of checklists to be completed. Resumption times (time to make the next move on the Tower of London task after the interruption) were longer in the 18-s interruption condition; however there was no effect for knowing the length of interruption in advance. This evidence from Hodgetts and Jones was the first empirical support of Altmann and Trafton's predictions for greater retrieval times for goal memories suspended for longer periods.

Despite Hodgetts and Jones's (2006b) findings, other studies failed to find a duration effect. In their landmark interruptions study, Gillie and Broadbent (1989) conducted a series of experiments investigating why some interruptions are more disruptive than others. In the first two experiments, participants performed a computer game task that required them to navigate through an environment and "pick-up" objects from a memorized list. Interruptions occurred after designated objects were picked-up and consisted of simple arithmetic problems. The first experiment used a 30-s interruption and the second experiment used a 2.75-min interruption; neither duration resulted in postinterruption task performance decrements. The authors claimed, "the length of an interruption on its own does not seem to be the critical factor in determining whether or not it will prove disruptive" (p. 246).

Recent research in the prospective memory domain provided additional data regarding the interruption duration question (Einstein et al., 2003; McDaniel, Einstein, Graham, & Rall, 2004). Einstein et al. found that people were able to maintain intentions over brief intervals ranging from 5 to 40 s. In a subsequent study, McDaniel et al. added a manipulation of interruption duration in the 40-s intention execution delay condition. They compared 10-s and 20-s interruption durations. The results once again showed no effect for intention execution delay, nor did they reveal an effect for interruption duration. McDaniel et al. argued that a maintenance rehearsal explanation should have resulted in a decline in prospective memory performance for the longer interruption; however, this prediction was not supported. More interesting, the digit monitoring interruption task used by Einstein et al. and McDaniel et al. probably reduced participants' ability to rehearse intentions during the interruption. These prospective memory findings were yet more evidence that contradicted the predictions of the memory for goals model.

The challenge for the Altmann and Trafton (2002) model was to explain why one of its fundamental predictions for interrupted task performance had not been supported in the interruptions literature until recently (Hodgetts & Jones, 2006b). A review of the literature revealed two reasons for the failure to find consistent evidence for an interruption duration effect. First, the measures used in many interruptions studies were global, and therefore insensitive to the effects associated with goal resumption. For example, Gillie and Broadbent (1989) compared pre- and postinterruption task times and error rates, which did not address the time participants required to resume the task after being interrupted. Czerwinski et al. (2000) measured total task time, the time to respond to the interruption notification, and the time spent on the interruption notification. Zijlstra et al. (1999) similarly measured task times and total interruption time, in addition to other performance measures. Other interruptions studies used measures like error rates in primary task performance (Cellier & Eyrolle, 1992; McFarlane, 2002; Oulasvirta & Saariluoma, 2004), decision-making performance (Speier et al., 1999; Speier, Vessey, & Valacich, 2003), and proportion of correct prospective memory responses (Einstein et al., 2003; McDaniel et al., 2004). The lack of sensitive measures for how quickly people resume the primary task after the interruption may have been one of the key reasons why previous studies failed to find an effect for interruption duration. It was not until Hodgetts and Jones (2006b) implemented Altmann and Trafton's (2002) resumption lag measure that evidence for the effect materialized. As a result, the resumption lag measure, which is a response time measure capturing the time required to resume a goal, was adopted for the present experiments. The intent was to capture the changes in resumption time using the resumption lag measure as predicted by the memory for goals model.

The second reason why past research failed to find a consistent effect for interruption duration was the manipulation of interruption duration. The interruption duration effect predicted by the memory for goals model (Altmann & Trafton, 2002) occurs when goals are still in the initial stages of decay. Although the 30-s and 2.75-min interruptions used by Gillie and Broadbent (1989) seemed reasonable in terms of face validity, these interruption durations may have masked resumption time effects. The memory for goals decay function (see Figure 1) indicates that the rate of decay slows down dramatically over time, and therefore if a goal had reached asymptotic levels of activation decay after 30 s, then the activation level for a goal suspended for more than 2 min would be similar, assuming the same level of initial activation. Therefore, the only way to detect the predicted effect was to use much shorter interruptions like the 6-s and 18-s interruptions used by Hodgetts and Jones (2006b).

Interruption Demand

Because the theoretical explanation for the duration effect focuses on goal memory decay, the issue of goal rehearsal must also be addressed. Goals left unrehearsed during an interruption will decay, resulting in longer resumption times (Altmann & Trafton, 2002; Hodgetts & Jones, 2006b). However, there are many interruption tasks that afford opportunities to rehearse the suspended goal. For example, the report writer could make quick glances to the document in the text editor when waiting for responses in the instant message exchange. These quick "reminder" glances to the report would help maintain the writer's train of thought for when the instant message conversion concludes. Alternatively, the writer may engage in an instant message conversation wherein a long, detailed response is made over several seconds, preventing any glances to the open report. In this scenario, the suspended thought would be difficult to resume without recreating the thought processes by reading the previous report entry. Therefore, it follows that interrupting tasks that prevent or inhibit goal maintenance should result in unmitigated goal decay manifested as longer resumption times. Alternatively, interrupting tasks that allow people to rehearse their suspended goals should show shorter resumption times in comparison (see Trafton et al., 2003).

The literature also suggested that ability to rehearse during an interruption is influenced by the cognitive demand of the interrupting task. Gillie and Broadbent (1989) showed that additional decoding requirements to an arithmetic task resulted in worse primary task performance. Zijlstra et al. (1999) found that document editing tasks resulted in more time to "reorient" to the primary editing task compared to interruptions consisting of unrelated menial tasks such as looking up a phone number. Recently, Cades et al. (2007) showed that 1-back and 3-back versions of the *n*-back task resulted in longer resumption times than a shadowing interruption task. Finally, Hodgetts and Jones (2006b) also manipulated interruption task complexity with a single digit addition task (simple) and a double-digit addition task requiring carrying (complex). They found that both the simple and complex interruptions resulted in longer resumption times than the no-task interruption condition. These studies provided evidence to support an effect for interruption complexity, both at the more sensitive resumption time measure and at the more global task measures as well.

Although the interruptions literature has generally supported an effect for interruption task complexity, the term *complexity* has been inconsistently defined. Gillie and Broadbent (1989); Hodgetts and Jones (2006b), and Cades et al. (2007) all used a processing requirements definition for complexity. These manipulations were consistent with the definition by Byrne and Bovair (1997), who noted that a number of characteristics appear to determine complexity, including the number of actions to be performed, the difficulty of executing those actions, the number of subgoals to be remembered, and the amount of information to be managed and maintained. For the purposes of this series of experiments, we used the term *demand* rather than complexity because it referred more directly to the processing demands on working memory that prevented or allowed rehearsal of suspended task goals.

Because the manipulation of rehearsal was complicated, it was assumed that the cognitive demand of an interruption task was directly related to the amount of resources available for rehearsal. In other words, more demanding interruption tasks would leave few resources, if any, for goal maintenance or rehearsal. Accordingly, working memory processing demands were varied in the interruption task in attempt to manipulate the available resources for rehearsal.

Overview of the Experiments

The present set of experiments was designed to test predictions from Altmann and Trafton's (2002) memory for goals model regarding the resumption of suspended memories (i.e., task goals). The fundamental prediction addressed by this set of experiments was that memory for task goals decays over time, resulting in longer resumption times for those task goals that have had more time to decay. The resumption lag measure introduced by Altmann and Trafton has shown to be sensitive to differences in task goal resumption times due to interruption complexity, duration, the interval between interruption alert and engagement (interruption lag), and cues in previous research (see Hodgetts & Jones, 2006a, 2006b; Monk et al, 2004; Trafton et al., 2003). Multiple and frequent interruptions within each trial characterized the interleaved tasks interruption paradigm used in this set of experiments. Although the majority of past interruptions studies used few interruptions per trial, the present focus was on interleaved interruptions similar to the instant messaging example. However, the tasks used in these experiments were not intended to simulate instant message interruptions. Instead, the intention was to use computer-based tasks that could be used to create situations in which interrupting tasks were interleaved with the primary task. In addition, the focus on resuming suspended task goals required a primary task with many subgoals that users typically perform linearly. A VCR programming task was selected because it has served this purpose well in past research (Monk et al., 2004).

The first experiment tested the prediction that longer interruptions lead to longer resumption times. The second experiment attempted to replicate the findings from Experiment 1 although extending the interruption duration manipulation to further characterize the decay trend for suspended goals. Finally, the third experiment added levels of task demand to the original interruption duration manipulation to test the rehearsal explanation of the duration effect.

Experiment 1

The first experiment was designed to test the predictions of Altmann and Trafton's (2002) model regarding time to resume suspended task goal memories using the interleaved tasks interruption paradigm. For the purposes of this experiment, goals were defined as low-level, next action goals. For example, the memory of what button to click next in a computer interface would be the suspended goal during an interruption or shift in attention. Consistent with the 6-s and 18-s interruption durations used by Hodgetts and Jones (2006b), Experiment 1 used interruption durations of 3, 8, and 13 s. In addition, Experiment 1 included uninterrupted control trials to assess the effect of interruptions on primary task performance. First, the general disruptiveness of interruptions was predicted to be evident in longer resumption lags compared to the average time between uninterrupted clicks (called *interaction intervals*). Second, the resumption times were predicted to increase from 3 to 13 s. Although the memory for goals model predicted a log function for resumption times over increased interruption durations, the segment of the function captured between 3 and 13 s was expected to appear linear.

Method

Participants

Twelve students from George Mason University received partial course credit for participating in this study. The participants (5 men and 7 women) ranged in age from 17 to 32, with an average age of 20 years.

Tasks and Equipment

The primary task was a VCR programming task using a simulated VCR built in Macintosh Common Lisp (Gray, 2000; Gray & Fu, 2001). The interruption task was a pursuit-tracking task that required participants to track a moving target. These tasks were presented side-by-side on a Macintosh G4 computer with a 17-inch VGA monitor. The VCR task was displayed on the left side of the monitor and the tracking task on the right side. Participants programmed show information into the VCR for randomly selected intervals between 3, 5, or 7 s at a time. The random VCR times were used to prevent participants from predicting the onset of interruptions. The VCR task was interrupted by the tracking task for 3, 8, or 13 s, alternating back and forth until the VCR program show was completed. Both tasks required only the computer mouse for input, and only one of the tasks was visible at a time.

VCR task. Programming a show consisted of four subtasks: entering the show's start time, end time, day of week, and channel number. The VCR interface can be seen in Figure 2. All interactions with the VCR were based on simulated VCR buttons; there were no field entries. To enter the start time, the participant first clicked the column button above the hour buttons (leftmost square button under the enter button). The participant then clicked the start-hour button, before clicking on the up or down arrow multiple times until the displayed hour number reached the target. Next, the participant clicked on the enter button to save the start-hour setting. Finally, to end this subtask, the participant clicked the column button again (to "deselect" it) before moving onto the next subtask. The same process was completed for each subtask element of the end time, day of week, and channel number tasks, respectively. The VCR display was blank when no setting was selected. The participants had access to target show information (the show name, start time, end time, day of week, and channel number) at all times as the information was posted to the right of the monitor on a 3 × 5-index card.

Interruption task. The pursuit-tracking task required the participant to track an airplane symbol (target) moving around the right half of the screen. The target's movement algorithm randomly updated each of the x and y coordinates by no more than 100 pixels (either direction) at a rate of 10 Hz. The resulting movement was rapid and somewhat erratic. The airplane symbol's visual angle was estimated at .37 degrees high by .79 degrees wide. The circle that corresponded to the participant's mouse position was estimated at .97 degrees of visual angle.

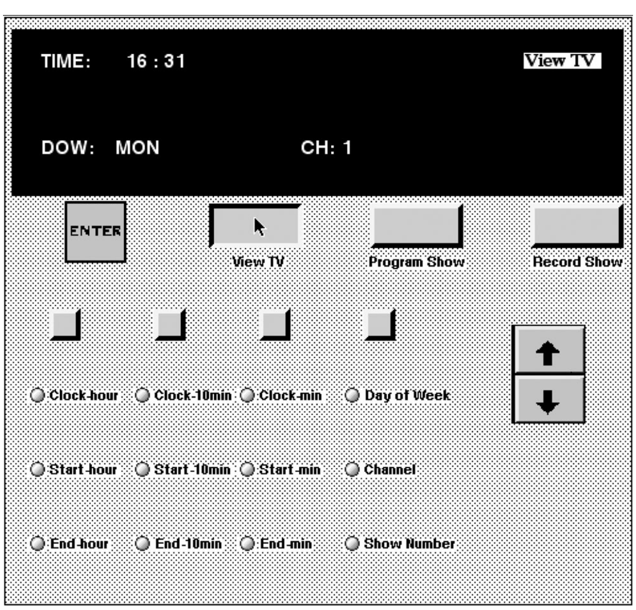

Figure 2. Simulated VCR interface used in primary task.

Design

A single factor repeated measures design was used to test the interruption duration hypothesis. There were three interruption durations, 3 s, 8 s, and 13 s, which were varied between trials. Each participant completed two trials for each interruption duration, resulting in 6 interruption trials. In addition, each participant completed 6 uninterrupted trials that served as a comparison for the interrupted trials for determining the magnitude of any disruption effect, for a total of 12 experimental trails. The dependent measure was the resumption lag after each interruption, as measured by the time from the switch from the tracking task to the VCR task until the participant's first click on a button in the interface. Participants tended to establish a consistent sequence of operations or "path" when programming the VCR. Therefore, resumption errors were identified as those clicks that deviated from the expected next action based on each participant's established path through the task. Tracking task and resumption error performances were also recorded to assess any potential speed–accuracy trade-off in performance. Trial order was randomized and balanced with a Latin square.

Procedure

Each participant was tested individually. The sessions, which lasted approximately 1 hr, began with the experimenter explaining the VCR task through demonstration. The participants were then given 2 practice trials in which they programmed the VCR without interruption, followed by two 60-s practice trials with the tracking task alone. The participants were then introduced to the interruption condition, in which they alternated performing the VCR and interruption tasks within a trial. The participants were instructed that the cursor position for each respective task would be repositioned to its saved location on each switch so that dragging the mouse back and forth between the two sides was unnecessary. Accordingly, the cursor position, along with various state indicators in the VCR interface (e.g., column button in "selected" state), acted as environmental cues that aided resumption. Participants were instructed to treat both tasks as equally important, and to focus on the task that was "on" at any given moment. Because the trials began and ended with the VCR task, there was implicit emphasis on this task as the primary task. After the two practice interruption trials, the participants completed the 12 experimental trials, each with new show information to be programmed. Participants began each trial with the VCR programming task. After completing the experimental trials, the participants were debriefed and dismissed.

Results and Discussion

The resumption lag data were screened for errors to isolate VCR task actions that represented successful postinterruption goal resumption. There were two categories of resumption errors. The first category consisted of resumption actions that deviated from the participant's established task path. Due to the nature of the VCR task, participants generally performed the task in the same sequence of actions across all trials. This reliable behavioral pattern provided a definition of path deviation for each participant. The second error category consisted of resumption lags less than

100 ms, which were assumed to be due to incidental mouse clicks timed coincidentally with the VCR task onset. Both resumption error categories were eliminated from the data. The path-deviation resumption error rate was 5.3%. Table 1 shows that the error rate was lowest in the 3-s condition, but consistent between the 8- and 13-s conditions. Repeated measures analysis of variance (ANOVA) revealed no significant difference between the error rates across the three conditions, $F(2, 22) = 1.17$, $p = .33$. There were no resumption lags less than 100 ms.

Data were lost for one participant's second 8-s interruption trial so five values from the population of 8-s trial resumption lags were randomly selected and imputed for the missing cell. The following results represent the mean values of the five calculations and analyses for each imputation (see McKnight, McKnight, Sidani, & Figueredo, 2006, for missing data procedures).

To first demonstrate the presence of a disruption effect for interruption trials, the resumption lags for the interrupted condition were compared with randomly sampled interaction intervals (IAIs) in the uninterrupted condition. The IAIs were the time elapsed between interface actions or button clicks, and were viewed as an appropriate comparison for the resumption lags to quantify the relative disruptive effect of interruptions in the VCR task. By comparing the mean resumption lags (M = 1,548 ms, SD = 231) to the mean IAIs for the uninterrupted trials (M = 949 ms, SD = 283), it was evident that the interruptions resulted in a delay in the execution time for the next action or goal in the interface compared to when uninterrupted. Subtracting the mean IAI from the mean resumption lag reveals an estimated cost of 599 ms on the VCR programming task. These data showed that resumption lags were longer than interaction intervals for uninterrupted trials, indicating the basic interruption disruption effect on task resumption with the interleaved tasks interruption paradigm.

To test the interruption duration prediction, the resumption lags from the interrupted trials were entered into a single-factor repeated measures ANOVA. Recall that the following is the mean F value of the five ANOVAs corresponding to each randomly imputed value as detailed in the previously described missing data procedures (McKnight et al., 2006). The main effect for interruption duration was significant, $F(2, 22) = 10.92$, $p < .01$, $\eta_p^2 = .50$. As can be seen in Figure 3, the resumption lags increased from 3-s to 8-s to 13-s interruptions ($b = 30.72$), thereby supporting Altmann and Trafton's goal decay predictions. This finding was consistent with the interruption duration results by Hodgetts and Jones (2006b) and therefore represented a significant addition to

Table 1
Resumption Error Rates

	Interruption duration											
	3 s		8 s		13 s		23 s		38 s		58 s	
Experiment	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	SD	*M*	*SD*	*M*	SD
1	.03	.18	.07	.25	.06	.23						
2	.02	.13	.06	.23	.05	.23	.09	.28	.10	.30	.10	.30
3												
No-task	.01	.12	.03	.16	.03	.16						
Tracking	.01	.10	.01	.12	.02	.13						
n-back	.07	.25	.05	.21	.05	.22						

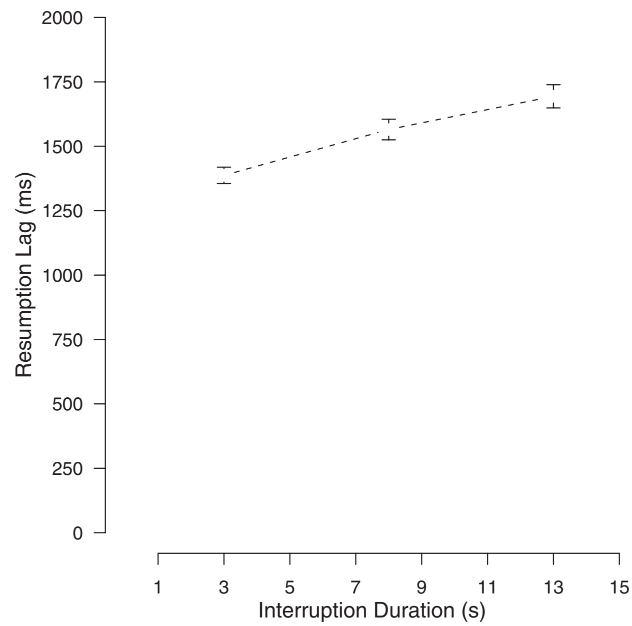

Figure 3. Mean resumption lags (±*SE*) as a function of interruption duration.

the growing body of empirical support for the memory for goals model as a framework for studying interrupted task performance (Altmann & Trafton, 2002; Trafton et al., 2003). *T* tests showed that the 3-s condition was reliably shorter than the 8-s and 13-s conditions (both $p < .01$), and that the 8-s condition was shorter than the 13-s condition ($p < .05$).

The *x-y* coordinates for the mouse and target positions were recorded at 10 Hz during the tracking task. The Euclidean distance between the target and mouse positions was calculated for each sampling record. The root mean square (RMS) of the distance calculations was used as a measure of accuracy. The first second of data (i.e., the first 10 data points) after each switch to the tracking task were excluded due to high variability although the participants readjusted to the tracking task. The RMS scores were averaged across the tracking task switches within trials, and again for participants within interruption duration conditions. The data were trimmed using a cut-off of three standard deviations above the mean. With this criterion, 2% of the tracking data were excluded. Tracking task performance was significant for the interruption duration, $F(2, 22) = 15.90$, $p < .01$, $\eta_p^2 = .59$. Table 2 shows that RMS was higher for the 3-s condition, though a Tukey's HSD post hoc analysis only revealed significant difference between the 3-s and 13-s conditions ($p < .05$). The worse performance in the 3-s condition was likely due to less time for tracking performance to stabilize compared to the longer durations.

Taken together, the relationship between interruption duration and resumption time from this experiment and the duration effect found by Hodgetts and Jones (2006b) provided strong evidence for the existence of an interruption duration effect despite the null findings of duration in previous interruptions studies (Gillie & Broadbent, 1989; Li et al., 2006; McDaniel et al., 2004). Without question part of the reason for the discrepancy was tied to the specific tasks in the different studies, but there were two reasons to

Table 2
Tracking Task Performance

	Interruption duration											
	3 s		8 s		13 s		23 s		38 s		58 s	
Experiment	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
1	47	11	41	9	40	8						
2	56	21	51	18	50	18	57	26	54	21	56	20
3	43	7	40	5	40	5						

Note. Given in root mean square.

suggest why both the present experiment and Hodgetts and Jones detected an effect of duration whereas past studies had not. The first reason was connected to using the resumption lag measure, which was appropriately sensitive to the theoretical predicted outcomes. The second reason was linked to the manipulation of interruption duration. Gillie and Broadbent were perhaps the least likely to detect an effect of interruption duration because they used longer durations (minimum 30 s) and global measures. McDaniel et al. manipulated delays in the shorter duration range (5 and 15 s), but they used global measures of intention execution. Li et al. found a trend for more postcompletion errors with a 45-s interruption compared to a 15-s interruption; however, this difference was not reliably different. Our results, combined with those from Hodgetts and Jones, suggested that the interruption duration effect was best detected with the resumption lag measure and durations less than 15 s.

An unanswered question was if the resumption lag trend would continue to increase linearly with longer interruption durations, or if the trend would resemble a log function as predicted by the memory for goals model. Absent additional strengthening (e.g., rehearsal) during an interruption, a suspended goal's activation level should decay as a function of delay (see Figure 1). As an indicator of goal memory activation, the resumption lag trend should be characterized as an inverse of the decay function, rapidly climbing in the shorter duration range (i.e., the duration effect) before approaching asymptote. Experiment 2 was designed to test this prediction.

Experiment 2

Experiment 2 was designed with two objectives in mind: To replicate the interruption duration effect in Experiment 1 and to extend the resumption lag profile beyond 13 s to nearly 1 min. Based on the Altmann and Trafton (2002) model, it was predicted that rate of resumption times would rise rapidly in the short duration range (i.e., 3 to 13 s) before the diminishing over the next 45 s, approaching asymptote (i.e., a log function). To meet these objectives, three additional longer interruption durations were added to the 3-, 8-, and 13-s interruptions used in Experiment 1. The longer durations were specified using increasing intervals of 10, 15, and 20 s, which resulted in 23-s, 38-s, and 58-s interruption durations. The increased variability in the interruption duration manipulation was intended to provide a resumption lag profile for durations ranging between 3 s and 1 min. We hypothesized that the resumption lag profile would best fit a log trend, resembling an inverse of the decay function (see Figure 1) as predicted by the memory for goals model.

Method

Participants

Twelve undergraduates from George Mason University received partial course credit for participating in this study. The participants (8 men and 4 women) ranged in age from 18 to 23, with an average age of 21 years.

Tasks and Equipment

The VCR and tracking tasks were identical to those used in Experiment 1.

Design

A single factor repeated measures design was used with six levels of interruption durations. The six durations were 3 s, 8 s, 13 s, 23 s, 38 s, and 58 s. There were no matched uninterrupted trials in this experiment. Participants completed 2 trials for each duration, resulting in a total of 12 experimental trials. The shorter (3 s, 8 s, and 13 s) and longer (23 s, 38 s, and 58 s) interruption duration trials were blocked and counterbalanced with a Latin square across participants.

Procedure

The procedure was identical to that in Experiment 1 except that all 12 experimental trials were interruption trials.

Results and Discussion

One participant's data were excluded from the analyses because of failure to perform the tracking task during the interruption. As with Experiment 1, both categories of resumption errors were removed from the data. The overall path-deviation resumption error rate was 7%. Table 1 shows that the error rate was again lowest in the 3-s condition and gradually increased with interruption duration. A repeated measures ANOVA revealed a significant difference between the error rates across the six conditions, $F(5, 50) = 3.55$, $p < .01$, $\eta_p^2 = .26$. Tukey's HSD post hoc comparisons showed that the 3-s condition was significantly lower than the 38-s and 58-s conditions. With the 100 ms criterion for incidental resumption actions, 0.3% of the resumption lag data were excluded from the analyses.

The purpose of this experiment was to show that resumption times followed a log function corresponding to activation decay over time. The memory activation formula as expressed in Altmann and Trafton (2002) was a log function that resulted in the familiar decay pattern (see Figure 1). Accordingly, a log model was fit to the data from this experiment. As seen in Figure 4, the model fit the means data very well ($R^2 = .989$). As interruption duration increased, the resulting resumption lag times grew at a slower rate. This finding was particularly important because it showed that the memory for goals model's explanation for interrupted task performance could account for not only the presence of the interruption duration effect at the shorter durations, but it also

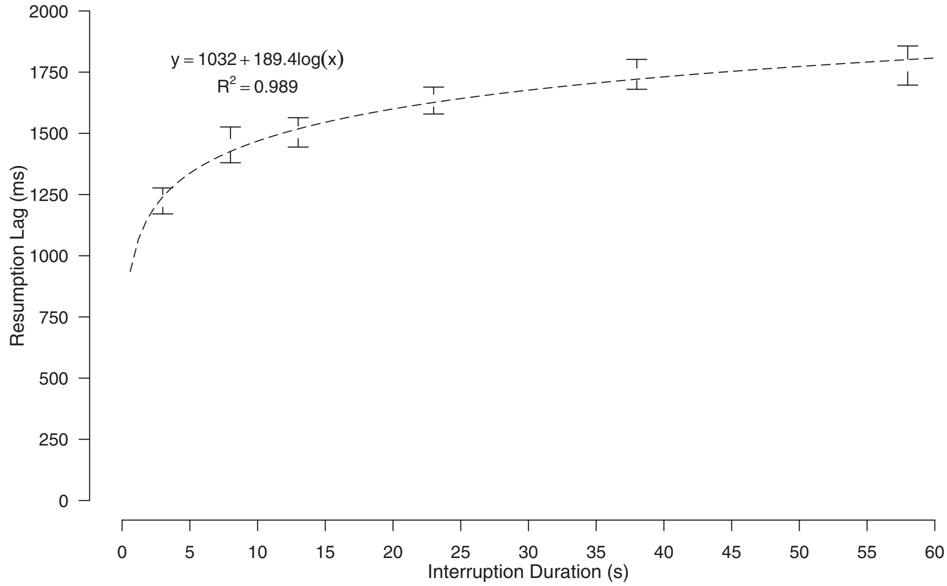

Figure 4. Mean resumption lags (±*SE*) as a function of interruption duration, with model fit.

offered an explanation regarding the absence of this effect in previous literature that used interruption times longer than 30 s (e.g., Gillie & Broadbent, 1989).

Although the model fit provided strong support for the theoretical prediction, there were two additional questions that required attention. First, the initial three durations were examined to determine if the duration effect from Experiment 1 was replicated within this range of durations. Second, we attempted to identify the range at which the resumption lag curve began to approach asymptote to provide an indicator of when interruption duration ceases to have substantial effect on resumption time.

As in Experiment 1, the 3-s, 8-s, and 13-s interruption durations resulted in a significant main effect for duration, $F(2, 20) = 11.95$, $p < .01$, $\eta_p^2 = .54$. Paired comparisons (*t* tests) showed that only the 3-s condition was reliably shorter than the 8-s and 13-s conditions (both *p*s $< .01$). The resumption lag slopes for the 3-s to 13-s durations were very similar for Experiments 1 and 2 ($b = 30.98$ and $b = 32.54$, respectively). Combined with the interruption duration findings of Hodgetts and Jones (2006b), the main effects for the 3-s to 13-s interruption duration conditions in Experiments 1 and 2 provided compelling support for the authenticity of the duration effect and the goal-activation explanation.

To identify if and when the resumption lag curve began to asymptote, the 58-s and 38-s conditions were compared and were found to not be significantly different, $t(21) = -.626$, $p = .54$. Moving one duration shorter, the linear contrast between the 23-s, 38-s, and 58-s conditions was marginally significant at best, $F(1, 10) = 3.81$, $p = .08$. Because the linear trend was close to being reliably different from zero, confidence was low in declaring the 23-s range as the point of asymptote. Accordingly, the 13-s condition was added to the analysis, which yielded a significant linear contrast for the 13-s through 58-s conditions, $F(1, 10) = 10.37$, $p < .01$. These results indicated that the resumption lag curve began to asymptote some time between 13 and 23 seconds for the VCR and tracking task pairing.

The tracking task performance data were trimmed and analyzed in the same manner as in Experiment 1, resulting in exclusion of 3% of the tracking data. Tracking task performance (see Table 2) failed to show a significant effect across the six interruption durations, $F < 1$.

The implication for interruptions and interleaved task situations is that brief interruptions will be less disruptive in terms of resuming the interrupted task, but only for interruptions lasting up to roughly 15 to 25 s when the effect appeared to approach asymptote. Conclusions beyond 1 min cannot be drawn from the present results, but they suggest that people desiring to interleave tasks should strive to shift attention at least every 15 s for optimal resumption times in computer-based, hierarchical tasks. Recall that Gonzalez and Mark (2004) found that information workers shifted between tasks every 3 min on average, and Isaacs et al. (2002) found that instant message turn taking occurred every 15 s on average.

Taken together, the results of Experiments 1 and 2 provided compelling evidence that the memory for goals framework can accurately describe the role of goal decay in interruption recovery. The model was further tested in Experiment 3, which focused on the interaction between interruption duration and varying levels of interruption task demand, which was assumed to be related to the ability of participants to engage in goal rehearsal during the interruptions. None of the recent interruptions studies based on Altmann and Trafton's (2002) memory for goals model manipulated both interruption duration and demand. Whereas Hodgetts and Jones (2006b) provided important empirical findings related to duration and demand, they did not manipulate these factors in the same experiment to test the interaction between the two.

Experiment 3

Because the memory for goals model explains the interruption duration effect in terms of memory activation decay, the role of rehearsal becomes important in understanding how people manage suspended goals. Theoretically, persistent rehearsal during an interruption, regardless of duration, should minimize the interruption duration effect. In other words, if the duration effect is due primarily to activation decay, then the rehearsal process of strengthening a memory's activation trace over the course of the interruption should make resuming that goal easier and faster, thereby minimizing the duration effect. Alternatively, if the ability to maintain the suspended goal through rehearsal is minimized with a demanding interruption task, then the duration effect may reveal higher rates of decay compared to the tracking task condition in which some rehearsal was assumed to be possible. Experiment 3 was designed to compare these three conditions to test the predictions of the memory for goals theory regarding the strengthening constraint and the interruption duration effect.

Three levels of interruption task demand were included to test the rehearsal prediction. These levels of demand were assumed to directly impact opportunity for goal rehearsal during the interruptions. Rehearsal was assumed to be uninhibited in the low-demand condition, moderately inhibited in the medium-demand condition, and severely inhibited in the high-demand task. For the low-demand condition, the interruption did not consist of a task. Rather, the interruption was a blank screen. Participants were free to rehearse goals during the no-task interruptions and it was assumed they would (see Trafton et al., 2003), although they were not specifically instructed to do so.

The medium-demand condition consisted of the tracking task used in Experiments 1 and 2. It was considered to be moderately demanding because of its largely perceptual-motor nature, which afforded opportunities for participants to rehearse their VCR task goal while tracking.

For the high-demand interruption condition, the selected task was a verbal version of the *n*-back task that required participants to listen to, remember, and make decisions about verbally presented letters. Although different from the verbal *n*-back task used by Smith and Jonides (1999), the same assumptions about executive processes and storage of verbal material applied. Pilot participants reported being unable to think about the VCR task while performing the verbal *n*-back task, suggesting that participants would have few remaining cognitive resources for rehearsal during the *n*-back task interruptions.

The interruption duration effect was predicted for both the tracking and *n*-back task conditions, with the latter demonstrating a steeper trend because of unmitigated goal decay. Alternatively, the uninhibited opportunity to rehearse in the no-task condition was predicted to minimize the duration effect as evidenced by a flatter slope than the tracking and *n*-back task conditions. Corollary predictions were that the mean resumption lags for the *n*-back condition would be longer than the other two conditions, and the resumption error rates would be highest in the *n*-back condition. In addition, it was predicted that the no-task condition would yield shorter resumption lags than the tracking task and *n*-back conditions because of uninhibited opportunities for goal rehearsal.

Method

Participants

Thirty-six undergraduates from George Mason University received partial course credit for participating in this study. The participants (9 men and 27 women) ranged in age from 18 to 30, with an average age of 21 years.

Tasks and Equipment

The VCR task was identical to those used in the previous experiments. The interruption task levels consisted of a no-task condition, the tracking task, and the *n*-back task. For the no-task condition, the interruption consisted of a blank screen. The tracking task was the same as in Experiments 1 and 2. A verbal version of the *n*-back task was used in the high-demand interruption condition.

The *n*-back task involves the serial presentation of digits where the participant must respond whether the current digit is higher or lower than the previously presented digit (e.g., Lovett, Daily, & Reder, 2000). For the verbal version of the *n*-back task, single letters were presented serially and participants are required to respond if the letter came before or after the 1-back letter in the alphabet. For this experiment, the letters were presented aloud by the computer and subjects responded by clicking on either a "higher" or "lower" button, corresponding to closer to Z or closer to A, respectively. For example, if the letter sequence was G followed by T the correct response was "higher." The letters were "spoken" by the computer at a rate of one letter every 1.6 s. The response buttons were located on the right half of the screen in place of the tracking task. As with the tracking task condition, the cursor was automatically repositioned to the saved position on the right or left half of the screen on a switch.

Design

This experiment was a 3 × 3 mixed within-between design. The 3-s, 8-s, and 13-s interruption durations were used as the within-subjects factor. The between-subjects factor was interruption demand, which included the no-task, tracking task, and *n*-back task conditions. Participants were randomly assigned to one of these three interruption demand conditions and performed six experimental trials, two for each level of interruption duration.

Procedure

The procedure was the same as in Experiment 1 with the exception that the no-task condition participants did not receive any interruption task practice, and the *n*-back task participants received two 60-s practice trials. In addition, the interruptions occurred in fixed 5-s intervals.

Results and Discussion

As with the previous experiments, path-deviation resumption errors and resumption lags less than 100 ms were screened from the data. The overall path-deviation resumption error rate was 3%. Table 1 presents the error rates for level of the interruption duration and demand factors. The error rate data were entered into a mixed within-between ANOVA. There was a significant main effect for interruption demand, $F(2, 27) = 10.56$, $p < .01$, $\eta_p^2 = .44$. Tukey's HSD post hoc comparisons revealed the error rate in the *n*-back condition ($M = .06$, $SD = .23$) was greater than the error rates in the no-task condition ($M = .02$, $SD = .15$) and the tracking task condition ($M = .01$, $SD = .12$), both $p < .01$.

The no-task and tracking task conditions were not reliably different ($p = .62$). Neither the main effect for duration nor the interaction between duration and demand was significant ($F < 1$). With the 100 ms criterion for incidental resumption actions, 0.3% of the resumption lag data were excluded from the analyses.

A 3 × 3 mixed within-between ANOVA revealed the predicted main effect for task condition, $F(2, 33) = 19.92$, $p < .01$, $\eta_p^2 = .55$. The *n*-back condition resulted in the longest resumption lags (M = 1,789 ms, SD = 340), followed by the tracking task condition (M = 1,605 ms, SD = 244), and then the no-task condition (M = 1,322 ms, SD = 239). Planned *t* test comparisons showed that each of these conditions was reliably different from the others (all $p < .01$). The fact that the three levels of interruption demand resulted in the predicted ordinal resumption lag outcome along with the greater resumption error rate in the *n*-back condition indicated that the task demand manipulation was a successful proxy for manipulating goal rehearsal opportunity. In addition, the faster resumption lags in the no-task condition supports the assumption that participants took advantage of the opportunity to rehearse.

The omnibus ANOVA also revealed a significant main effect for interruption duration, $F(2, 66) = 19.96$, $p < .01$, $\eta_p^2 = .38$. However, the predictions concerned the individual demand conditions rather than the overall main effect. To determine if the interruption duration effect was present within demand condition, separate repeated measures ANOVAs were conducted for the no-task, tracking, and *n*-back task conditions. The no-task condition resulted in a significant main effect for duration, $F(2, 22) = 4.57$, $p < .05$, $\eta_p^2 = .29$. In contrast with the findings from Experiments 1 and 2, the main effect for duration was not significant in the tracking task condition, $F(2, 22) = 2.08$, $p = .15$, $\eta_p^2 = .16$. Finally the *n*-back task condition resulted in the predicted main effect for duration, $F(2, 22) = 17.24$, $p < .01$, $\eta_p^2 = .61$.

The absence of the duration effect in the tracking condition was surprising given its reliability in the previous experiments. However, the lower effect size compared to those in Experiments 1 and 2 (.50 and .54, respectively) suggested that the lack of effect might have been due to greater variability in participants. Further examination of the tracking condition results revealed a marginally significant linear trend, $F(1, 11) = 4.62$, $p = .055$, $\eta_p^2 = .30$, hinting of the duration effect. Considering the overall main effect for duration combined with the effects and trends at the task demand condition level, there was compelling evidence to accept the duration effect despite its modest presence in the tracking condition. The presence of the duration effect in the no-task condition suggested that despite uninhibited opportunity for goal rehearsal, goal activation still showed evidence of decay as interruption durations increased.

The significant interaction between interruption duration and demand conditions, $F(4, 66) = 3.92$, $p < .01$, $\eta_p^2 = .19$, was also of interest because the *n*-back and tracking conditions were predicted to produce steeper duration effect trends than the no-task condition. Specifically, it was hypothesized that the *n*-back task condition would result in a steeper trend than the tracking task and no-task conditions because of limited available cognitive resources while performing the cognitively demanding *n*-back task, and that both the *n*-back and tracking task conditions would product steeper slopes than the no-task condition. Linear contrast interactions were conducted between the three demand conditions to test this hypothesis.

As seen in Figure 5, the linear contrast interaction between the *n*-back and the no-task conditions was significant, $F(1, 22) = 11.17$, $p < .01$, $\eta_p^2 = .48$, as was the interaction between the *n*-back and tracking task conditions, $F(1, 22) = 8.26$, $p < .05$, $\eta_p^2 = .27$ as predicted. These differences were confirmed with an analysis of the slopes. The slope for the *n*-back condition (b = 39.14) was significantly greater than the slope for the no-task condition (b = 11.39), $t(11) = -3.61$, $p < .01$. The *n*-back slope was also greater than the tracking task condition slope (b = 12.60), $t(11) = -2.80$, $p < .05$, as predicted. However, the no-task and tracking task condition slopes were not significantly different, $t < 1$. The linear contrast interaction between the tracking and no-task conditions was also not significant, $F < 1$. Whereas the *n*-back task produced a greater resumption lag slope than the tracking and no-task conditions as predicted, the tracking task condition slope was much lower than it was in Experiments 1 and 2 (b = 30.98 and b = 32.54, respectively). The marginal duration effect in the tracking task condition, as evidenced by the smaller slope, suggested that the slope interactions with the tracking task condition be considered cautiously.

The tracking task performance data were analyzed as in the previous experiments, along with the *n*-back task performance. The tracking task performance data were trimmed and analyzed in the same manner as in the previous experiments, resulting in exclusion of 1% of the tracking data. Tracking task performance (see Table 2) did not vary reliably across the three interruption durations, $F(2, 22) = 2.90$, $p = .076$. The *n*-back task accuracy scores were computed for each trial. Because the letter presentation rate (every 1.6 s) prevented a response to the second stimulus in the 3-s condition, related no-response errors were screened out of the data. Accuracy rates showed no difference between the three

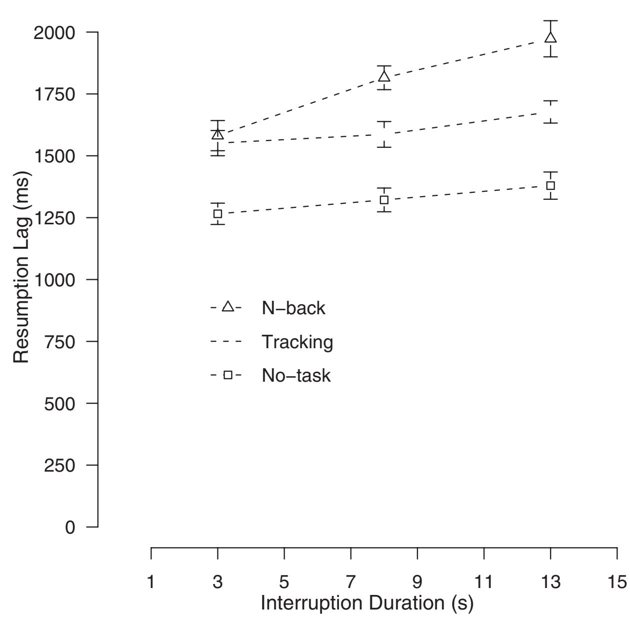

Figure 5. Mean resumption lags (±*SE*) as a function of interruption duration and demand.

interruption durations, $F < 1$. The mean accuracy rate was 73% ($SD = 12\%$) for the 3-s condition, 71% ($SD = 7\%$) for the 8-s condition, and 74% ($SD = 4\%$) for the 13-s condition.

It is important to note the consistent performance in the *n*-back and tracking tasks across the three interruption durations. Combined with no differences in the resumption error rates across duration, there was no evidence of a speed–accuracy trade-off to explain the interruption duration effect. The large differences in resumption times between task demand conditions support the view that rehearsal is key to efficient resumption of suspended goals in an interruption situation (see Trafton et al., 2003). However, the presence of the duration effect in the no-task condition suggested that even in optimal rehearsal conditions, decay processes appeared to win out to some degree.

General Discussion

The goal of this set of experiments was to demonstrate that interruption duration and demand affect postinterruption task resumption, and that goal decay and opportunity to rehearse play an important role in these effects. Experiment 1 showed that interruptions were disruptive, and that longer interruptions were associated with longer resumption times, as predicted by the memory for goals model (Altmann & Trafton, 2002). This finding was consistent with the duration effect between 6-s and 18-s interruptions demonstrated by Hodgetts and Jones (2006b). Experiment 2 extended the interruption duration manipulation to nearly 1 min and supported the predicted log function for resumption lags. The trends observed in Experiments 1 and 3 were not inconsistent with the log trend observed in Experiment 2 because the 3-s to 13-s segment of the resumption curve captured the steep incline period that resembles a linear trend. Finally, Experiment 3 showed that resumption lags were longer when available resources for rehearsal were minimized by a high-demand interruption task. The results also showed an interaction between duration and demand manifested by a steeper resumption lag trend across durations in the high-demand condition. The combined evidence from all three experiments supported the veracity of the interruption duration effect, its log function characteristic over interruptions up to 1 min, and highlighted the importance of goal rehearsal during interruptions for better resumption performance. These findings will be discussed in terms of their theoretical and practical implications.

Theoretical Implications

The results of this study contributed to the growing body of interruptions literature in terms of the effects of interruption duration and cognitive demand, and how these two factors interact to impact resumption of suspended task goals. The findings indicated that the time to resume a task after an interruption depended both on the duration of that interruption and the cognitive demand of the interrupting task. We argued that the duration effect was primarily due to goal memory decay, and that the demand effect was directly related to the ability to rehearse the suspended goal during the interruption. Each of these factors was found to affect resumption performance by Hodgetts and Jones (2006b), and the present studies confirmed and expanded on their interruption duration findings to create a resumption lag profile from 3 to 58 s. In addition, unlike Hodgetts and Jones's study, Experiment 3 manipulated both duration and demand to show how interruption task demand impacts the duration effect. Specifically, the results showed that a more cognitively demanding interruption task produced a steeper resumption lag trend across the 3-s, 8-s, and 13-s durations. This interaction showed for the first time how opportunities to rehearse not only helped to speed-up resumption times, but also showed how rehearsal opportunities help mitigate goal memory decay as interruption duration increases. This finding highlighted the importance of interleaving quick "reminders" of the primary task state for reducing resumption costs.

In the instant messaging example, the report writer could occasionally steal a glance to the report or quickly think about the suspended goal while waiting for a quick response, or even before reading a response. In other words, people can interleave rehearsal within just about any task that does not consume the available cognitive resources (see Trafton et al., 2003). The presence of the decay trend in the no-task condition strongly suggested that despite optimal rehearsal opportunities, decay effects still manifest in resumption times.

Quality of goal rehearsal may be partly responsible for the slight decay trend in the no-task condition. A mismatch between the type of rehearsal executed and the actual task goal could have weakened the strengthening of the task goal (see Nairne, 2002). Another possibility was that the type of rehearsal that people engaged in was somehow shallow or ineffective for maintaining activation levels above the interference threshold. Einstein et al. (2003) attempted to deal with this issue by instructing participants to use implementation intentions as a means of having participants form detailed plans for accomplishing intentions after a delay. Implementation intentions are the more detailed when, where, and how aspects of accomplishing the goal intention (Gollwitzer, 1999) rather than the intention to accomplish a goal. Einstein et al. predicted better prospective memory performance by instructing participants to form implementation intentions rather than simple goal intentions. The assumption was that by generating implementation intentions the participants would be encoding more detailed and therefore stronger intentions. However, the implementation intentions proved no better than simple rehearsal instructions for remembering to execute an intention over brief delays. Further research is required to fully explore the rehearsal characteristics that produce optimal goal strengthening in memory.

Goal decay is an important component of the memory for goals model (Altmann & Trafton, 2002) and is consistent with the findings from classic short-term memory studies such as Brown (1958) and Peterson and Peterson (1959), which showed longer retention intervals led to more forgetting. The present study provided strong support for the role of decay in the memory for goals over short interruptions. However, a common criticism of the decay process of forgetting is that interference can be used to explain the same effects. The interference that occurs during the interruption may better explain our findings rather than the time-based process of memory decay and goal rehearsal. Perhaps people were more likely to experience proactive interference as time away from the primary task increased because of the build-up of previous task goals in memory. Recent evidence suggested that intrusion errors were greater for an interrupted task, but the intrusions were based on prior-knowledge rather than on the interruption itself (Oulasvirta & Saariluoma, 2004).

Contrary to the proactive interference explanation, Monk (2004) found that resumption lags were actually shorter when people were interrupted more frequently. More frequent interruptions should result in greater proactive interference because more goals have been suspended and resumed. Monk suggested that the rapid switching between the VCR and tracking tasks may have compelled participants to adopt a strategy to actively rehearse their suspended goals during the interruptions, leading to faster resumption times. Whether this find was viewed as lack of evidence for proactive interference using the same empirical method or as evidence for the active rehearsal strategy explanation, the results were consistent with the memory for goal model's decay explanation. In addition, Altmann and Schunn (2002) made a compelling argument for the role of decay in short-term forgetting. They did not argue that decay is the principal mechanism for forgetting; rather that it plays a secondary but important role compared to interference. Likewise, the importance of interference as a strong contributor to forgetting is not disputed here; however, the present evidence shows that goal decay also plays an important role.

The results from interruption studies are inevitably compared to those from task-switching studies in which switch costs have been explored extensively (see Monsell, 2003, for a brief review). This comparison is particularly tempting with the interleaved interruptions paradigm from the current study. However, interruption studies involve the suspension and resumption of task goals rather than the switching of stimulus-response mappings between trials, which we argue is a fundamentally different operation. Hodgetts and Jones (2006b) noted that time-based determinants of goal retrieval cannot be attributed to task-switching costs, and that the memory for goals model provided a more compelling explanation for resumption costs. Mixing cost evidence from the task-switching literature (see Monsell, 2003; Rubin & Meiran, 2005), however, provided an alternative theoretical explanation for the duration effect that was important to consider. Mixing costs are those costs associated with maintaining multiple task sets in working memory, resulting in longer response latencies in switching trials versus single-task trials. The duration effect, therefore, could have been the result of different resource allocation strategies when maintaining two task goals in memory in the shorter verses longer interruption durations.

In the present experiments, mixing costs would translate to longer IAIs in the interrupted versus uninterrupted VCR programming trials in Experiment 1. Knowing that they would need to interleave the VCR and tracking tasks, participants may have maintained both task sets in working memory to foster better switching performance. The differential allocation of resources would be an overall effort-saving strategy to produce more efficient task switching and thus better dual-task performance overall when interleaving two tasks. However, the dual-task strategy loses its utility with longer interruptions because the switches seem few and far between (though there were actually the same number of switches on average because the VCR times consistently ranged between 3 and 7 s). The strategy changes to exclusively allocate resources to the tracking task until the shift back to the VCR task. The change to single-task resource allocation would result in longer resumption times when switching back to the VCR task because the VCR task set was not actively maintained during the interruption.

When comparing the IAIs from the interrupted and uninterrupted trials from Experiment 1, we found the opposite results from those predicted by the resource allocation explanation. The IAIs in the interrupted condition averaged 510 ms (SD = 81), whereas they averaged 949 ms (SD = 284) in the uninterrupted condition (using the same sampling procedure as in Experiment 1). However, this finding did not rule out the resource allocation explanation entirely because Rubin and Meiran (2005) showed that mixing costs were eliminated when the two task sets were unambiguous (i.e., clearly distinct tasks) as they were in this study.

If the resource allocation explanation was correct, then we would have expected to see consistently short resumption lags until the interruption duration was sufficiently long to elicit the strategy shift, producing longer, asymptotic resumption lags. One would expect to see a resumption lag trend resembling a logistic *s*-curve across the six durations in Experiment 2 rather than the observed log function (see Figure 5). As long as people were working to maintain both tasks sets in working memory, faster resumption lags should have resulted. However, once the dual-task strategy was abandoned for the single-task strategy, one would expect asymptotic resumption lags. In fact, the observed resumption lag trend from Experiment 2 supported the goal decay explanation over the resource allocation explanation.

The present findings also added to a growing body of empirical evidence (e.g., Altmann & Trafton, 2007; Cades et al., 2007; Hodgetts & Jones, 2006a, 2006b; Li et al., 2006; Monk et al., 2004; Trafton et al., 2003) supporting the use of the memory for goals model (Altmann & Trafton, 2002) as a framework for studying interruptions. When the interruptions problem was approached with this cognitive theory, we were able to make specific predictions that were confirmed by using a theory-driven metric that is sensitive to the subtle effects of goal decay. Corroborating and extending Hodgetts and Jones' (2006b) duration effect evidence while also demonstrating why this effect has gone undetected in previous research (e.g., Gillie & Broadbent, 1989) was indeed a powerful expression of how sound cognitive theory can significantly contribute to the interruptions problem.

One issue that was unaddressed by this research was the role of environmental cues in helping to retrieve suspended goals. In the VCR task, there were several available cues to help the participant reestablish the suspended task state. For example, the cursor arrow remained in the same location as when the switch occurred, providing a powerful cue as to where the participant was in the task and what action/goal was to be accomplished next. Other display features such as button activation highlights and display feedback were also available to aid the participant in resuming the task. However, these cues were available in all conditions and still the interruption duration and inhibited rehearsal effects persisted. More research is required to fully isolate the role of environmental cues from rehearsal, recency, and frequency.

Practical Implications

The application of these findings to real-world tasks exceeds the simple conclusion that longer and more demanding interruptions will result in longer primary task resumption times. We discuss some of the contexts in which people interleave interrupting tasks, and where additional time costs when shifting cognitive effort can have significant ramifications. In addition, we discuss how the duration and demand findings generalize to each of these situations.

Quick switches between the primary tasks and interrupting tasks, or task interleaving, is a common behavior observed in contexts such as driving, emergency rooms, and aviation cockpits, among others. Studies describing glance duration and frequency behavior when engaging an in-vehicle tasks go back decades (e.g., Dingus, Antin, Hulse, & Wierwille, 1989; Mourant & Rockwell, 1972; Wierwille, 1993). The results from these studies and others showed that voluntary eyes-off-road times rarely exceed 2 s. Wierwille argued that tasks requiring more than 1.5 s to complete push drivers to adopt a time-sharing strategy shifting visual and cognitive attention between the driving and in-vehicle tasks. Gellatly and Kleiss (2000) showed that people were remarkably consistent in shifting attention between the road and in-vehicle task every second. This time-sharing scenario showed how people interleave tasks in a similar time scale as studied in the present experiments. The resumption costs in the present experiments were on the order of hundreds of milliseconds. The time costs certainly were inextricably connected to the VCR programming task used in this study, along with the tracking and *n*-back interruption tasks; however, Lee, Caven, Haake, and Brown (2001) showed that reaction latencies as short as 300 ms can greatly increase the odds of a collision. Therefore, quick shifts of attention can potentially have consequences in both driver reactions to unexpected events as well as time to complete the in-vehicle task. The longer a task takes to complete, the more time the driver spends engaged in a distracting and potentially dangerous situation. The connection between driver reactions and resumption costs should be considered cautiously until further research using driving tasks and resumption lag measures are conducted.

The interruption duration effect has less of an impact on the driver distraction situation because of the small range of observed glance durations (see Horrey & Wickens, 2007). However, the demand effect does have implications for the kinds of tasks that drivers engage in while driving. Our results suggest that simple tasks such as tuning the radio dial (visual and motor requirements only) would have lower time costs than a complex task like finding a particular song in an Mp3 player or entering a destination into a GPS navigation system (visual, cognitive, and motor requirements). Because our findings rely on the resumption of suspended task goals, generalizing to reactions to driving-related events that do not involve goal resumption should be made with caution. Further research is required to quantify switch costs on driver reactions. Alternatively, our findings help to understand total time to complete a task like destination entry because a task goal must be suspended and resumed with each shift of attention. Even if the resumption lags were very short for each shift, the costs would be additive over the course of the entire task, resulting in longer task times. Longer task times are typically associated with more eyes-off-road time because attention must be shifted a greater number of times.

Emergency rooms are another environment in which people shift visual and cognitive attention frequently and rapidly. Chisholm et al. (2001) reported that emergency room physicians spent 37.5 min out of every hour managing three or more patients and were interrupted 9.7 times per hour. Although it is impossible to estimate from our data how resumption costs manifest in emergency rooms, our data showed that repeated suspension and resumption of task goals may be costly in such a time-critical context. For example, an alarm may sound during a procedure requiring several seconds of a nurse's attention. Once the urgent matter is resolved, the nurse then returns to the previous task of assisting the doctor, potentially with a brief time delay as the nurse retrieves the suspended goal from memory. As in the driving example, the demand effect has the potential to be greater than the duration effect because of the range of cognitive tasks in such complex, life and death situations. Further research investigating resumption performance in emergency rooms and other health care environments is crucial for understanding how interruptions affect performance and ultimately patients' lives.

Another critical situation in which interruptions can have a significant impact is the aircraft cockpit. Air traffic controllers, other personnel in the cockpit, and flight attendants frequently interrupt pilots going through preflight checklists and other critical tasks. Loukopoulous, Dismukes, and Barshi (2001) reported that frequent interruptions in the cockpit required pilots to continuously make task management decisions, including adding, shedding, and rescheduling actions. Perhaps more important than the time costs associated with interruptions in the cockpit are the potential error costs such as missed items on the preflight checklist. As previously noted, multiple-plane crashes have been attributed in part to interruptions to the preflight checklists (NTSB, 1969, 1988).

There are countless other situations in which people interleave tasks. The instant messaging example was used earlier to show how interruption duration and demand could affect the resumption of a writer's performance. This example does not typically involve life-threatening situations as with the driving and emergency room examples; however, the additive time costs can have significant economic impact in loss of productivity over time.

Conclusions

The goal of this study was to apply a well-specified theory of memory for goals to the real-world problem of resuming tasks after being interrupted. The results helped define the role of interruption duration and demand in determining resumption costs. Duration was shown to result in increased resumption costs when the interruptions lasted between 3 and 13 s; however, a log function pattern emerged when the duration manipulation was extended to nearly one minute. This finding supported the role of decay in Altmann and Trafton's (2002) theory. Demand was also shown to have a substantial impact on resumption costs, indicating that opportunities to rehearse suspended task goals are an important determinant in defining resumption times. The interaction between duration and demand, although needing further exploration, provided additional insight into how opportunity to rehearse task goals during an interruption can help mitigate decay processes, though it appeared that decay cannot be completely eliminated even with optimal opportunity for goal rehearsal. These results added to the growing body of empirical support for the memory for goals model and its application to the study of interruptions (see Altmann & Trafton, 2007; Cades et al., 2007; Hodgetts & Jones, 2006a, 2006b; Li et al., 2006; Monk et al., 2004; Trafton et al., 2003). The current findings also provided insight into the practical costs of interleaving interruption tasks with the primary task. The added resumption times associated with interruptions have important consequences for overall task efficiency and productivity in office settings; however, these costs can have far more serious consequences in situations like driving, emergency rooms, and aircraft cockpits.

References

Altmann, E. M., & Schunn, C. D. (2002). Integrating decay and interference: A new look at an old interaction. *Proceedings of the 24th annual meeting of the Cognitive Science Society* (pp. 65–70). Mahwah, NJ: Erlbaum.

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

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

Anderson, J. R. (1993). *Rules of the mind.* Hillsdale, NJ: Erlbaum.

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review, 111,* 1036–1060.

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

Avrahami, D., & Hudson, S. E. (2006). Responsiveness in instant messaging: Predictive models supporting inter-personal communication. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'06*) (pp. 731–740). New York: ACM.

Brown, J. (1958). Some tests of the decay theory of immediate memory. *Quarterly Journal of Experimental Psychology, 10,* 12–21.

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

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

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

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

Chisholm, C. D., Dornfeld, A. M., Nelson, D. R., & Cordell, W. H. (2001). Work interrupted: A comparison of workplace interruptions in emergency departments and primary care offices. *Annals of Emergency Medicine, 38,* 146–151.

Cutrell, E., Czerwinski, M., & Horvitz, E. (2001). Notification, disruption and memory: Effects of messaging interruptions on memory and performance. In *Human-Computer Interaction–Interact '01* (pp. 263–269). Amsterdam: IOS.

Czerwinski, M., Chrisman, S., & Rudisill, M. (1991). *Interruptions in multitasking situations: The effects of similarity and warning* (Tech. Rep. No. JSC–24757). Houston, TX: NASA Johnson Space Center.

Czerwinski, M., Cutrell, E., & Horvitz, E. (2000). Instant messaging: Effects of relevance and time. In S. Turner & P. Turner (Eds.), *People and computers 14: Proceedings of HCI* 2000 (pp. 71–76). London British Computer Society.

Czerwinski, M., Horvitz, E., & Wilhite, S. (2004). A diary study of task switching and interruptions. In *CHI 2004 Proceedings* (pp. 175–182). New York: ACM.

Dingus, T., Antin, J., Hulse, M., & Wierwille, W. (1989). Attentional demand requirements of an automobile moving-map navigation system. *Transportation Research, 23*, 301–315.

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

Einstein, G. O., McDaniel, M. A., Williford, C. L., Pagan, J. L., & Dismukes, R. K. (2003). Forgetting of intentions in demanding situations is rapid. *Journal of Experimental Psychology: Applied, 9,* 147–162.

Gellatly, A., & Kleiss, J. (2000). Visual attention demand evaluation of conventional and multifunction in-vehicle information systems. *In Proceedings of the XIVth Triennial Congress of the International Ergonomics Association and 44th Annual Meeting of the Human Factors and Ergomonics Society* (pp. 3.282–3.285). Santa Monica, CA: Human Factors and Ergonomics Society.

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

Gollwitzer, P. M. (1999). Implementation intentions: Strong effects of simple plans. *American Psychologist, 54,* 493–503.

Gonzalez, V., & Mark, G. (2004). "Constant, constant, multi-tasking craziness": Managing multiple working spheres. *Proceedings of ACM CHI'04* (pp. 113–120). New York: ACM.

Gray, W. D. (2000). The nature and processing of errors in interactive behavior. *Cognitive Science, 24,* 205–248.

Gray, W. D., & Fu, W.-T. (2001). Ignoring perfect knowledge in-the-world for imperfect knowledge in-the-head: Implications of rational analysis for interface design. *CHI Letters, 3,* 112–119.

Hodgetts, H. M., & Jones, D. M. (2006a). Contextual cues aid recovery from interruption: The role of associative activation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32,* 1120–1132.

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

Horrey, W. J., & Wickens, C. D. (2007). In-vehicle glance duration. *Transportation Research Record No. 2018,* 22–28.

Iqbal, S. T., & Horvitz, E. (2007). Disruption and recovery of computing tasks: Field study, analysis, and directions. In *Proceedings of CHI* 2007 (pp. 677–686). New York: ACM.

Isaacs, E., Walendowski, A., Whittaker, S., Schiano, D. J., & Kamm, C. (2002). The character, functions, and styles of instant messaging in the workplace. *Proceedings of the ACM Conference on Computer Supported Cooperative Work* (pp. 11–20). New York: ACM.

Kreifeldt, J. G., & McCarthy, M. E. (1981). Interruption as a test of the user-computer interface. In *Proceedings of the 17th Annual Conference on Manual Control (Publication 81–95*) (pp. 655–667). Pasadena California Institute of Technology, Jet Propulsion Laboratory.

Lee, J. D., Caven, B., Haake, S., & Brown, T. L. (2001). Speech-based interaction with in-vehicle computers: The effect of speech-based e-mail on drivers' attention to the roadway. *Human Factors, 43,* 631–664.

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

Loukopoulos, L. D., Dismukes, R. K., & Barshi, I. (2001). Cockpit interruptions and distractions: A line observation study. *The 11th International Symposium on Aviation Psychology.* Columbus: Ohio State University.

Lovett, M. C., Daily, L. Z., & Reder, L. M. (2000). A source activation theory of working memory: Cross-task prediction of performance in ACT–R. *Journal of Cognitive Systems Research, 1,* 99–118.

Lovett, M. C., Reder, L. M., & Lebiere, C. (1999). Modeling working memory in a unified architecture: An ACT–R perspective. In A. Miyake & P. Shah (Eds.), *Models of working memory* (pp. 135–182). Cambridge, MA: Cambridge.

McDaniel, M. A., Einstein, G. O., Graham, T., & Rall, E. (2004). Delaying execution time of intentions: Overcoming the costs of interruptions. *Applied Cognitive Psychology, 18,* 533–547.

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

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

McKnight, P. E., McKnight, K. M., Sidani, S., & Figueredo, A. J. (2007). *Missing data: A gentle introduction.* New York: Guilford.

Monk, C. A. (2004). The effect of frequent versus infrequent interruptions on primary task resumption. In the *Proceedings of the 48th Annual Meeting of the Human Factors and Ergonomics Society* (pp. 295–299). Santa Monica, CA: HFES.

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

Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences, 7,* 134–140.

Mourant, R. R., & Rockwell, T. H. (1972). Strategies of visual search by novice and experienced drivers. *Human Factors, 14,* 325–335.

Nairne, J. S. (2002). Remembering over the short-term: The case against the standard model. *Annual Review of Psychology, 53,* 53–81.

Newell, A. (1990). *Unified theories of cognition.* Cambridge, MA: Harvard University Press.

National Transportation Safety Board. (1969). *Aircraft accident report: Pan American World Airways, Inc., Boeing 707–321C, N799PA, Elmendorf Air Force Base, Anchorage, AK, December 26, 1968* (NTSB/AAR–69/08). Washington, DC: Author.

National Transportation Safety Board. (1988). *Aircraft accident report: Northwest Airlines, Inc., McDonnell Douglas DC-9–82, N312RC, Detroit Metropolitan Wayne County Airport, Romulus, MI, August 16, 1987* (NTSB/AAR–88/05). Washington, DC: Author.

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

Peterson, L. R., & Peterson, M. J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology, 58,* 193–198.

Rubin, O., & Meiran, N. (2005). On the origins of the task mixing cost in the cuing task-switching paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 1477–1491.

Smith, E. E., & Jonides, J. (1999). Storage and executive processes in the frontal lobes. *Science, 283,* 1657–1661.

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

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

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

Wierwille W. W. (1993). Visual and manual demands of in-car controls and displays. *In B. Peacock & Karwowski (Eds.), Automotive ergonomics* (pp. 299–320). London: Taylor and Francis.

Zeigarnik, B. (1927). Das behalten erledigter und unerledigter handlungen (The retention of completed and uncompleted actions). *Psychologische Forschung, 9,* 1–85.

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

Received October 2, 2007
Revision received September 17, 2008
Accepted September 25, 2008 ■