LEARNING TO MORE EFFECTIVELY MANAGE INTERRUPTIONS OVER
REPEATED EXPOSURES: WHEN, HOW, AND WHY?

by

Erik Thomas Nelson
A Dissertation
Submitted to the
Graduate Faculty
of
George Mason University
in Partial Fulfillment of
The Requirements for the Degree
of
Doctor of Philosophy
Psychology

Committee:

_____ Director

_____

_____

_____ Department Chairperson

_____ Program Director

_____ Dean, College of Humanities
and Social Sciences

Date: _____ Spring Semester 2013
George Mason University
Fairfax, VA

Learning to More Effectively Manage Interruptions Over Repeated Exposures: When, How, and Why?

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at George Mason University

by

Erik Thomas Nelson
Master of Arts
George Mason University, 2010
Bachelor of Science
University of Kansas, 2008

Director: Deborah A. Boehm-Davis, University Professor and Associate Dean
Department of Psychology

Spring Semester 2013
George Mason University
Fairfax, VA

DEDICATION

I dedicate this dissertation to my loving wife Adrienne, who has been amazingly supportive throughout these five years. I couldn't have done this without you!

To my son, William, who unknowingly gave me that added sense of urgency to get this dissertation finished.

To my parents, Tom and Lynn, who have always been supportive in everything I do. I wouldn't be where I am today without you.

To my sister, Kayla, who through friendly sibling rivalry, has always pushed me to work harder and be better at everything I do.

To Patrick Berry, who introduced me to Human Factors, and has been helping to guide me at each major professional milestone.

Finally, to my lifelong friend, Justin Martinek, who has always wanted me to mention him in one of my scholarly works.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF FIGURES

ABSTRACT

LEARNING TO MORE EFFECTIVELY MANAGE INTERRUPTIONS OVER
REPEATED EXPOSURES: WHEN, HOW, AND WHY?

Erik Thomas Nelson, Ph.D.

George Mason University, 2013

Dissertation Director: Dr. Deborah A. Boehm-Davis

Previous research on interrupted task performance has focused heavily on factors that

affect the resumption lag, the time it takes to resume from an interruption to the main task

(Altmann & Trafton, 2002; Hodgetts & Jones, 2003,2006). More recently, several studies

have found that the resumption lag tends to get shorter over repeated exposures – even

after accounting for general task learning (Trafton, Altmann, Brock & Mintz, 2003;

Cades, Boehm-Davis, Trafton & Monk, 2011). The purpose of this set of experiments

was to evaluate the ability of the three main theories of interrupted task performance

(Memory for Goals, Threaded Cognition, and Long Term Working Memory) to explain

the reductions seen in resumption lag over repeated exposures. Experiment 1 examined

the shape of the resumption lag distribution over repeated exposures as each theory

suggested a different distribution. Findings suggested that the mechanism proposed by

Long Term Working Memory was unlikely to be causing the reduction in resumption lag.

Experiment 2 tested a specific mechanism postulated by the theory of Threaded

Cognition (problem state). The results suggested that the reduction in resumption lag was not likely the result of the problem state mechanism. Experiment 3 examined the speeded encoding mechanism derived from Memory for Goals; the results indicated that improved ability to encode while performing the interruption task was not driving the reduction in resumption lag over repeated exposures to interruptions. Together, this set of experiments does not definitively support any single theory; however, the data do suggest which mechanisms are not likely to be driving the reductions seen in resumption lag over time, and suggest future experiments that may be able to determine the mechanisms underlying the reductions in resumption lag.

INTRODUCTION

Interruptions are pervasive in just about everything that we do. Studies with office workers have found that interruptions occur as often as 4-6 times per hour (Czerwinski, Horvitz, & Wilhite, 2004). Office workers aren't the only ones affected; research has also found interruptions to be prevalent in high risk environments such as healthcare (Flynn et al., 1999), aviation (Latorella, 1998) and nuclear power generation (Bainbridge, 1984). Unfortunately, not only are interruptions prevalent, they are also costly. Interruptions in the workplace have been calculated to cost the American economy 588 Billion dollars a year in lost productivity (Basex, 2006). The costs of interruptions in high risk environments such as healthcare, aviation and nuclear power generation are not only financial, but can also lead to injury and death. Interruptions have been identified as being responsible for as much as 43% of all medication administration errors (Santell, 2005) and have been found to be key contributors to aviation (Latorella, 1998) and nuclear incidents (Bainbridge, 1984).

Due to the prevalent nature of interruptions, they have been heavily researched. Much of this research has been devoted to the factors that affect performance when a task is interrupted. These include task-based and environmental factors, such as interruption complexity, duration, similarity to the interrupted task, interruption lag and point of interruption, as well as individual differences, such as strategies, training, experience, and

current workload. Although this research has been essential in helping us begin to understand why interruptions are disruptive, it is still unclear how and in what situations people get better at being interrupted across repeated exposures.

The purpose of this series of studies is to expand upon the limited research that has focused on the reduction of resumption lag, the time between finishing an interruption and resuming the primary task, over repeated exposures. Past research has shown that resumption lags tend to get shorter with repeated exposures to the same interruption task pair. However, it is not clear why or how this occurs. There are three major theories of interrupted task performance which provide potential mechanisms for this increase in performance over repeated exposures. In the following experiments, these theories' explanations will be tested to determine what theory's explanation is driving this effect.

In order to explore the question of why interruptions are disruptive, common terminology must be understood. Figure 1 presents Trafton et al.'s (2003)"anatomy of an interruption."
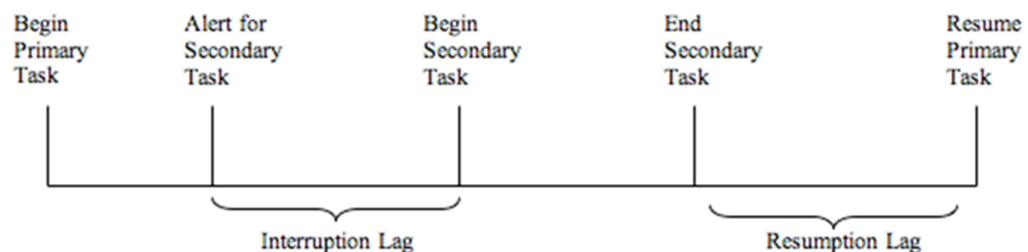
Figure 1 The anatomy of an interruption (borrowed from Trafton et al., 2003)

The timeline begins with the start of the primary task. In some cases an interruption (secondary task) is preceded by an alert, while in other situations an interruption has no warning. In the case where an interruption alert is present, an interruption lag can be calculated. It is calculated as the time between the onset of the interruption warning and the interruption itself. The presence of nonzero interruption lags have been shown to help improve interrupted task performance (Hodgetts & Jones, 2003). The next point on the timeline is the end of the interruption (secondary) task. Following the end of the secondary (interruption) task, there is a period of inaction followed by resumption of the primary task. As can be seen in Figure 1, this period of inaction is called the resumption lag. The resumption lag is the most common measure of interrupted task performance because of its high level of sensitivity (Altmann & Trafton, 2002).

Theories of Interrupted Task Performance

Although considerable research has been conducted on interrupted task performance, relatively few theories have been formulated that accurately predict how and under what circumstances a person will or will not be successful in resuming from an interruption. Below, the three main theories are discussed.

Memory for Goals

One of the most widely accepted theories of interrupted task performance is Memory for Goals (Altmann & Trafton, 2002). Memory for Goals is based on ACT-R, a computational modeling framework (Anderson et al., 2004; Anderson & Lebiere, 1998). In the ACT-R model, activation results from two sources: base level activation and

strength of association. Base level activation refers to how easily someone is able to retrieve a piece of information. The more a piece of information is activated, the easier it becomes to recall. This type of activation increases with experience. The other type of activation in the ACT-R model, strength of association, results from the current context. The current context is largely driven by the availability of both mental and environmental cues. If many cues are available, then the goal will be more highly activated due to the strength of association.

In the original ACT-R model, goals had a privileged status such that they never decayed over periods of inaction. The Memory for Goals model modified the basic ACT-R model, allowing goals to decay over periods of inaction. Further, the Memory for Goals model claims that goals are what drive behavior, and the goal with the highest level of activation tends to be the determinant of action. When a goal is highly activated and continues to be activated, the same goal will persist unless it is interrupted. However, once interrupted, the activation of a goal is suspended and the goal will decay until it is activated again. Therefore, the more time spent away from a goal, the more the goal will decay, and the more difficult it will be to resume.

Once another goal is more highly activated than the original goal, the newer, highly activated, goal will now drive action. Although it is beneficial to have relevant goals highly activated, it is important that they are not so highly activated as to cause proactive interference for future goals. This is called the interference level constraint, and it represents latent memories for old goals. The theory states that the more goals have been activated in the past, the more "noise" there is in the system, and the more difficult

it is to activate any one goal to criterion – the point at which goal-directed action may begin.

The Memory for Goals model is particularly helpful at predicting interrupted task performance. According to the model, interruptions activate a new goal unrelated to the original task. As this new goal increases in activation, it begins to drive action. Meanwhile, if the original goal is not being actively maintained through a strategy such as rehearsal, it will continue to lose activation. Once the activation of the original goal falls below threshold, it can no longer be recalled in an efficient manner. The Memory for Goals theory states that activation of this goal may only be increased by two different methods: priming and strengthening (Altmann & Trafton, 2002). During the process of strengthening, the Memory for Goals theory postulates that mental resources are focused on sampling memory for the most active goal. This quickly strengthens the activation of the primary task goal above the interference level caused by retroactive interference of past goals. Finally, once the primary task goal has been sufficiently strengthened, it may begin to drive action. Although strengthening can keep a goal from decaying, once the goal has fallen below the interference threshold, Memory for Goals states that priming through the use of cues is the only way to revive that goal, other than by chance (Altmann & Trafton, 2002). That is, Memory for Goals states that unless the primary task goal is primed with a related cue, it will have a difficult time overcoming retroactive interference. These cues can be any task-relevant feature that is present at the time of the interruption, as well as when the primary task is trying to be resumed (Hodgetts & Jones, 2006).

According to Memory for Goals, resumption lags may get shorter over repeated exposures due to improved goal encoding or increases in base level activation. Current research has found that the reduction in resumption lag is only seen after immediate interruptions, and not after interruptions with an interruption lag (Trafton et al., 2003). Therefore, if improved encoding is driving this reduction in resumption lag, then it may be due to increased ability to encode while concurrently completing the interruption task. The other probable mechanism driving the reduction in resumption lag is increased base level activation. The base level activation is composed of history and context-based activation. The more times a specific goal is recalled, the higher level of history-based activation it will have, thus increasing its base level activation. Memory for Goals would suggest that this higher level of activation would make it easier to recallthat specific goal, possibly leading to a reduction in resumption lag.

Long Term Working Memory

Another theory of interrupted task performance is Long Term Working Memory (LTWM). This theory states that if a person is an expert at a task, he can quickly code task-relevant information to generic retrieval structures located in what they call long term working memory (Ericsson & Kintsch, 1995; Oulasvirta & Saariluoma, 2004). Once this information has been coded into long term working memory, it is no longer subject to forgetting. If a person is not an expert at a task, LTWM theory suggests that he will still be able to code task-relevant information into long term working memory, however, this process will take much longer than if he were an expert. With respect to interrupted task performance, LTWM theory states that because experts are able to very quickly code

information directly into long term memory, they are not subject to forgetting, regardless of how long or difficult the interruption may be. However, LTWM theory does not predict it is impossible for experts to forget. Ericsson & Kintsch (1995) explain that the conversion to LTWM does not happen instantaneously, and if experts are interrupted before this transfer process is complete, the interruption will negatively affect the expert. They do not, however, provide specific times that it takes to encode to LTWM. Instead, they note that the duration of the encoding process varies based on the task and level of expertise.

According to the long term working memory theory, participants should be able to resume the primary task more quickly after they have become an expert compared to when they were not an expert. Although this theory does provide a possible explanation as to why there is a reduction in resumption lag over time, it may be a stretch to claim that participants have reached expert level performance after completing a task for 20-30 minutes. However, because some primary tasks are not terribly complex, it may be possible that participants could achieve "expert" level performance during the course of the experiment. If this effect is what is driving the reduction in resumption lag over time, then we would expect resumption lag to remain fairly constant until the transfer into long term working memory can happen fast enough to occur before the start of the secondary task. At this point, we would expect resumption lag to drop quickly and then level out at the lower level.

Threaded Cognition

A more recent model of interrupted task performance was proposed by Salvucci & Taatgen (2008). This model, Threaded Cognition, attempts to unite findings within multitasking, task switching, and interruptions. It describes all three areas as being on different points of a multitasking continuum and it bases this continuum roughly on the amount of time spent on one task before switching to another. At one end is multitasking as it is traditionally defined – completing one or more tasks concurrently. Task switching is at an intermediate point on the continuum, where the tasks are now too difficult to complete concurrently, so one must switch back and forth in order to complete them in a threaded, but serial manner. Finally, interruptions are on the far end of the continuum. Although similar to task switching, the time course of an interruption tends to be somewhat longer and there is not an intention to return back to the interruption after it has been dealt with.

Threaded Cognition is based on the ACT-R architecture and it incorporates many of the Memory for Goals assumptions, extending the work of Altmann & Trafton (2002). Salvucci & Taatgen (2008) describe Memory for Goals as an important first step to understanding the suspension and resumption of task goals. Meanwhile, they consider Threaded Cognition more of an in depth, step-by-step process model of the interruption-resumption process. One major contribution that they add is the concept of task threads. In the traditional model, a person suspends the primary task, completes the secondary task, and then switches back to the primary task with some degree of lag between the end of the interruption task and resumption of the primary task (Figure 2.a). Threaded

cognition allows for a rehearsal process where two threads are active at the same time. These threads follow a greedy-polite structure – greedy because they will use any available resource that is available to further their goal, but polite because they will return all resources as soon as they are finished using them. This structure allows for very efficient resource sharing, allowing several threads to run concurrently if they aren't using the same resources.
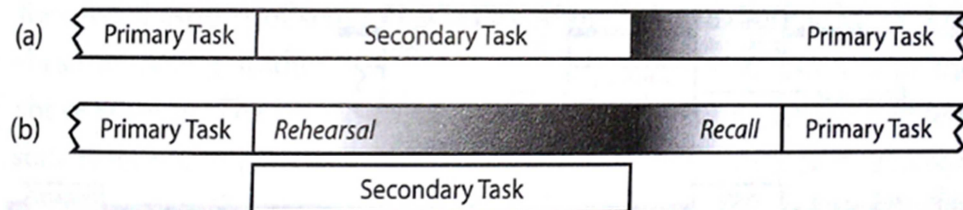
Figure 2 Task threads of the traditional interruption/resumption process (a) and Threaded Cognition's account of the interruption/resumption process (b) (borrowed from Salvucci & Taatgen, 2011)

Figure 2.b shows how the primary task is the primary task thread, and the interruption is the secondary task thread. While completing the primary task by itself, only the primary task thread is active. However, once the primary task is interrupted, the secondary task thread becomes active and the primary task thread continues to remain active. The secondary task thread immediately begins to process the interruption task while the primary task thread begins rehearsing the primary task in order to allow for successful resumption.

Another major contribution of Threaded Cognition to interruptions research is the concept of a problem state resource. Although Salvucci and colleagues did not introduce the problem state resource (Anderson, Qin, Sohn, Stenger, & Carter, 2003), they were the first to apply it to sequential multitasking and interruptions research. Problem state is defined as a resource that maintains information pertinent to accomplishing a task that is not easily accessible in the environment. Neuroimaging studies (Anderson et al., 2004; Anderson, 2007) have identified problem state as a distinctly different resource than the maintenance of goal information. Salvucci and Taatgen simplify these findings by explaining that "what a person is doing (the goal) and the information for doing it (the problem state) are maintained in distinct areas of the brain" (Salvucci & Taatgen, 2010, p. 114). Salvucci & Taatgen suggest that multiple goals may be active at any one time, but not multiple problem states. According to the model, a problem state rehearsal process occurs during the very first part of the interruption, and consists of a simple repetition of problem state retrievals until either n seconds have passed, or the problem state can be recalled within n seconds. If both the primary task and interrupting task have a problem state, one's ability to rehearse the primary task's problem state during the interruption is compromised, which can lead to increased time to recall the primary task as well as impaired performance for both the primary and interruption tasks.

In an effort to gather empirical evidence for the integration of the problem state resource into the Threaded Cognition model, Borst, Taatgen and van Rijn (2010) conducted an experiment manipulating the need to maintain multiple problem states. This experiment used a task switching paradigm where participants switched back and forth

between a subtraction task and a typing task. For each task, there was an easy version that did not require maintenance of a problem state, and a hard version that did require maintenance of a problem state. The easy version of the subtraction task did not require borrowing while the hard version required borrowing six out of ten times per task. For the typing task, the easy version showed the 10-letter word to be typed the entire time, while in the hard version the 10-letter word to be typed was only briefly displayed. These two conditions were fully crossed such that all participants experienced all four possible combinations of tasks. As expected, the easy-easy condition resulted in the fastest task switches with the highest level of accuracy, the easy-hard conditions resulted in an intermediate level of performance, and the hard-hard condition resulted in the longest task switches and lowest accuracy rates. Interestingly, there was a significant interaction where the difference between the easy-easy condition and the easy-hard conditions was quite small while the difference between the easy-hard and hard-hard conditions was rather large. This interaction supports the Threaded Cognition hypothesis that it is very difficult or even impossible to maintain two problem states (as evidenced by the hard-hard condition) at the same time.

However, Borst et al. (2010) conceded that there are other possible explanations for their data. An alternative hypothesis could be that participants are maintaining the problem state in the phonological loop and that the increased workload is causing the above interaction. In order to refute this alternative hypothesis, Borst et al. repeated the experiment, but this time a task switch occurred after every two similar trials rather than after every one trial. According to Threaded Cognition, the same interaction effect should

be seen during switch trials, but there should be no difference between conditions for non-switch trials because the problem state has not changed. According to the alternative workload explanation, similar differences should be seen during both switch and non-switch trials because workload should be constant. The results supported Threaded Cognition, as there was no difference between conditions in non-switch trials. This series of studies provide the first evidence that multiple threads cannot effectively maintain two separate problem states.

The main mechanism that Threaded Cognition suggests could lead to reduced resumption involves the problem state resource. Salvucci & Taatgen (2010) hypothesize that some procedures can be learned so well that declarative memory is no longer needed to recall where and how to resume after an interruption. Instead, procedural memory takes over. The advantage of procedural memory is that it does not require a problem state. Thus, a task may at first require a problem state, but later no longer require a problem state after sufficient exposure to the task. Salvucci & Taatgen hypothesize that this process may be responsible for reductions in resumption lag over repeated exposures.

Understanding How Resuming is Learned Over Time

Two studies have explored the reduction in resumption lag over time. In their study, Trafton, Altmann, Brock & Mintz (2003) set out to better understand how the interruption lag, the time between an interruption alert and the interruption itself, affects one's ability to resume from an interruption. To investigate this, they manipulated the presence of an interruption lag between subjects. One condition had an eight-second interruption lag and the other condition had no interruption lag – an immediate

interruption. Each participant completed three 20-minute sessions of a resource allocation

task where their goal was to command a group of tanks and destroy enemy targets. In

each session, participants were interrupted 10 times with a tactical assessment task which

consisted of classifying objects as either hostile or neutral. They found that the

introduction of an interruption alert significantly reduced resumption lag compared to an

interruption without an alert. However this effect became smaller across the three

sessions. Although the resumption lag for the alert condition seemed to remain relatively

low, the interruption lag for the immediate condition started relatively high and got

progressively faster across the three sessions (see Figure 3).
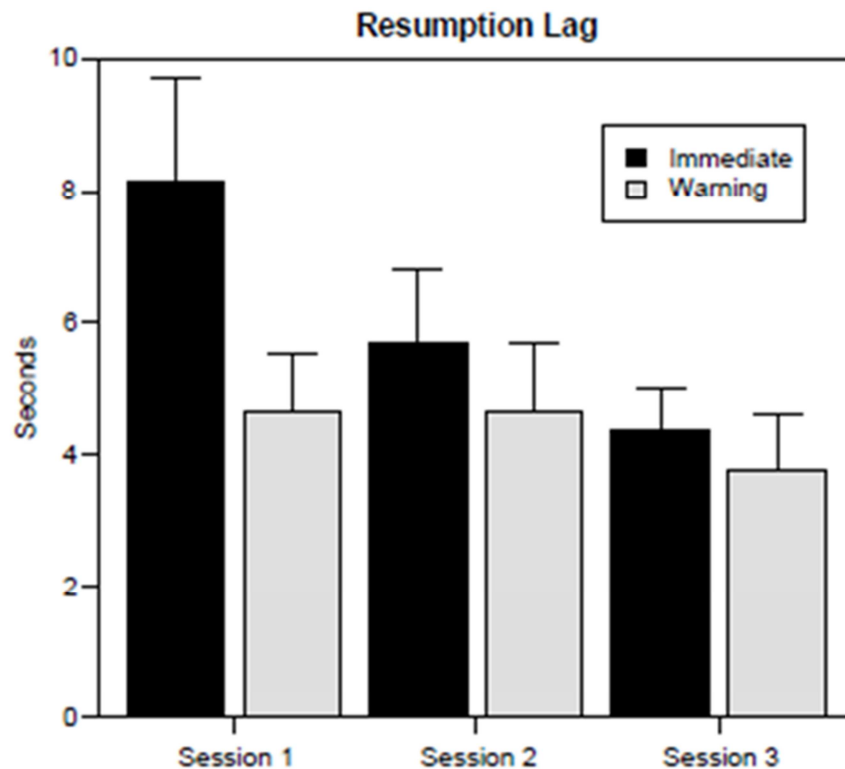
Figure 3 Main findings of Trafton et al. (2003)

These findings not only have implications for the effectiveness of interruption lags, they also suggest the possibility of a learning effect, and provide some insight into the overall interruption/resumption process. First, the data show that an interruption alert can be used to reduce the resumption lag. This provides solid evidence that participants effectively used the interruption lag to either better encode the primary task or find a better stopping point. Next, the immediate warning condition led to a consistent decrease in resumption lag across repeated exposures. Although far from conclusive, these data provide evidence of a learning effect for interrupted task performance. Finally, and most interestingly, while the immediate condition saw a consistent reduction in resumption lag

across repeated exposures, the alert condition remained consistently low with no reduction in resumption lag. The only difference between the two conditions was that the interruption lag provided the opportunity to encode the primary task and prepare for an interruption, and the immediate condition did not. This may suggest that there was something about one or both of these processes that got faster in the immediate condition while the alarm remained consistently low. Because interruptions occurred randomly and without warning in the immediate condition, it is unlikely that participants had any time to prepare. Therefore, the most likely process driving this reduction in resumption lag over time is better or more efficient encoding.

Even though Trafton et al. (2003) provided evidence of a learning effect, it was still not clear how or under which circumstances participants were able to resume from interruptions more quickly. Therefore, Cades, Boehm-Davis, Monk and Trafton (2011) extended this work in order to better understand these questions. In their study, they used the same primary and interrupting tasks as Trafton et al. (2003). Additionally, they also had participants complete 3 sessions, each with 12 random interruptions. In the first experiment, they used immediate interruptions and manipulated the amount of practice that participants had on both the primary task by itself (no interruption) and the primary task being interrupted by the secondary task (interruption). They had three conditions which manipulated what participants experienced across sessions. In the first condition, participants experienced no interruptions across all three sessions. In the second condition, participants were not interrupted during the first session, but were then interrupted in sessions two and three. In the final condition, participants were interrupted

in all three sessions. They found that resumption lag decreased with practice, but only when participants had been interrupted in the prior session.

In the final Cades et al. (2011) experiment, the authors set out to determine whether the learning effect was due to general experience with interruptions, or whether learning was specific to primary task/interruption task pairs. In this experiment, they used a different primary task, called the VCR task, which required participants to program the recording of a television show on a simulated VCR interface. The interruption task consisted of either a pursuit-tracking task or a shadowing task requiring the participant to repeat a displayed number out loud. In this experiment, participants were interrupted during all three sessions, however, the type of interruption was manipulated. In condition 1, participants had the same interruption all three sessions (AAA). In condition 2, participants had the same interruption the first two sessions and then a different interruption in session three (AAB). In condition 3, participants experienced one interruption type in the first session and then the other interruption type in sessions two and three (ABB). If the learning effect was due to general exposure to interruptions, all three conditions should see a consistent reduction in resumption lag across the three sessions. However, if the learning effect is interruption task-pair specific, then a reduction in resumption lag would only be expected between sessions one and two for condition 3, between sessions two and three for condition 2, and across all sessions of condition 1. They found that resumption lag decreased with exposure of two back to back interruptions of the same type. However, when the interruption task changed, resumption lags stayed the same, or even got worse. This provided strong evidence that the learning

effect is not due to a process specific mechanism such as encoding or resuming from an interruption. Instead, the learning effect appears to depend on a primary and interruption task specific interruption/resumption process.

Although these two experiments were able to confirm some kind of learning process where people get faster at resuming from interruptions with repeated exposures, it is still not clear what mechanisms are driving this phenomenon. Cades et al. (2011) were able to determine that people do not get better at being interrupted in general. Instead, reductions in resumption lag were only seen after practicing the same primary task-interruption task pair. This provides more information about the nature of the phenomenon, but not what is driving the phenomenon. The purpose of this series of experiments is to better understand how the reduction in resumption lag occurs.

EXPERIMENT 1

Rationale

All prior studies examining the reduction in resumption lag over repeated exposures have averaged resumption lag across experimental session when doing their analyses. In these prior studies, participants completed three experimental sessions with each experimental session containing approximately 10-12 interruptions (Trafton et al., 2003; Cades et al., 2011). The average resumption lag for each of these sessions was then compared to one another using an analysis of variance. These analyses were sufficient to determine if the resumption lag was getting shorter over time, however they are insufficient to test the more fine-grained features of the current theories of interrupted task performance.

Although this has not been considered in prior studies, the shape of the distribution of resumption lags over time can provide evidence for or against the three theories discussed above. These theories offer differing explanations for the mechanism behind the reduction in resumption lag over time. Moreover, the mechanisms behind each of these explanations would suggest different patterns of reduction in resumption lag over repeated exposures. Thus, the goal of Experiment 1 is to

As described above, the Memory for Goals and Threaded Cognition theories share several potential explanations for the reduction in resumption lag over repeated

exposures. The first explanation is that encoding of the primary task while completing the secondary task occurs more efficiently over repeated exposures. If this mechanism is driving the reduction in resumption lag over repeated exposures, we would expect an exponential or linear reduction in resumption lag over time (A or B in Figure 4, respectively). Two distributions are suggested instead of one because the shape of the distribution will depend on how quickly the learning process occurs. A Boltzmann sigmoidal model, where resumption lag remains relatively constant over time, followed by a sharp drop over a short period of time, followed by another period of consistent resumption lag (Figure 4.C) would not be expected.

The other explanation for the reduction in resumption lag shared by Memory for Goals and Threaded Cognition is increased base level activation. This mechanism suggests that the more the primary task is interrupted and maintained, the higher its base level activation becomes, which makes it easier to resume upon being interrupted. If the reduction in resumption lag were due to increased base level activation, we would expect an exponential reduction or possibly a linear reduction in resumption lag (Patterns A or possibly B in Figure 4, respectively). However, the increase in base level activation is likely to happen more quickly, due to the rapid rise in task specific activation, making pattern A the most likely candidate. As seen in Figure 4, all three mechanisms shared by Memory for Goals and Threaded Cognition suggest a gradual reduction in resumption lag, not an all or none phenomenon, as depicted by pattern C.

	A	B	C
Memory for Goals Encoding	YES	YES	NO
Base level Act.	YES	POSSIBLE	NO
Threaded Cognition Problem State	NO	YES	POSSIBLE
Long term working memory	NO	NO	YES
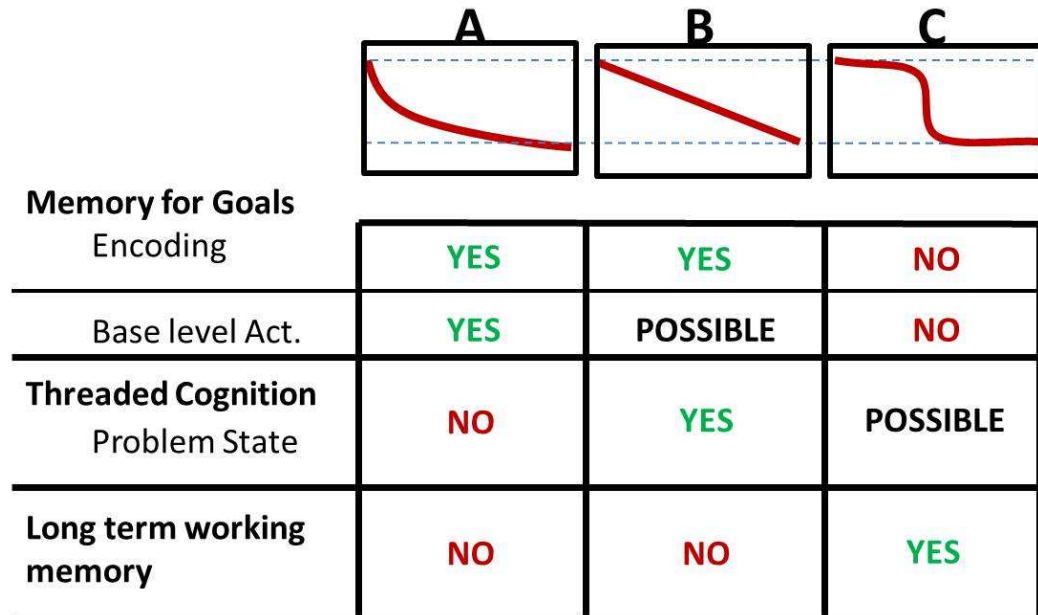
Figure 4 Patterns in the reduction in resumption lag as predicted by Memory for Goals, long term working memory, and Threaded Cognition theories.

Although Threaded Cognition is based on ACT-R and Memory for Goals, the problem state resource is the mechanism that Salvucci & Taatgen (2010) explicitly use to explain the reduction in resumption lag across repeated exposures. They state that when first completing a procedural task, knowledge on how to complete the task is stored in declarative memory. However, after a person has become familiar with the procedural task, the knowledge of how to complete the task moves from declarative to procedural memory. Threaded Cognition theory states that because declarative memory requires a problem state and procedural memory does not, it should be easier and faster to resume when using procedural memory. Therefore, the pattern of reduction could either take the

20

shape of pattern B or C, depending on how gradually or suddenly this transition from declarative to procedural memory occurs.

The long term working memory explanation for the reduction in resumption lag also has implications for its pattern of reduction. LTWM theory suggests that as a person has become an "expert," systems of retrieval cues are created which are referred to as retrieval structures. These "structures" are created in long term memory and allow information to be encoded very quickly and without the decay or forgetting associated with working memory. As long as the information can be encoded before needing to begin an interruption task, a person should be able to resume the primary task very quickly. Long Term Working Memory theory suggests that resumption lag performance will not change until a person can effectively use this specialized structure and encode the necessary details of the primary task fast enough before beginning the interruption task. Therefore, LTWM theory would suggest that the distribution of resumption lags would fit a Boltzmann sigmoidal model, where resumption lag remains relatively constant until this structure can be effectively used; at which point, a sharp decrease in resumption lag will occur, followed by a relatively constant resumption lag. This distribution corresponds with pattern C in Figure 4.

Method

Tasks

The purpose of this study was to better understand the mechanism behind the reduction in resumption lag across repeated exposures to interruptions. To accomplish this, the task had to be well controlled and easily repeatable. Therefore, the primary task

was a video cassette recorder (VCR) programming task written in Macintosh Common Lisp (Gray, 2000) – the same task used in Cades et al. (2011). Figure 5 shows the interface of the VCR task.

In order to successfully program the VCR to record a television show, a participant was required to set a show's start time, end time, day of week, and channel number. Before the beginning of the experiment, participants were given 3 x 5 inch index cards pinned to the side of the monitor with the show information required to program the VCR. Upon starting, participants had to first click the "View TV" button to "deactivate" the television function and then click the "program show" button to "activate" the programming function.
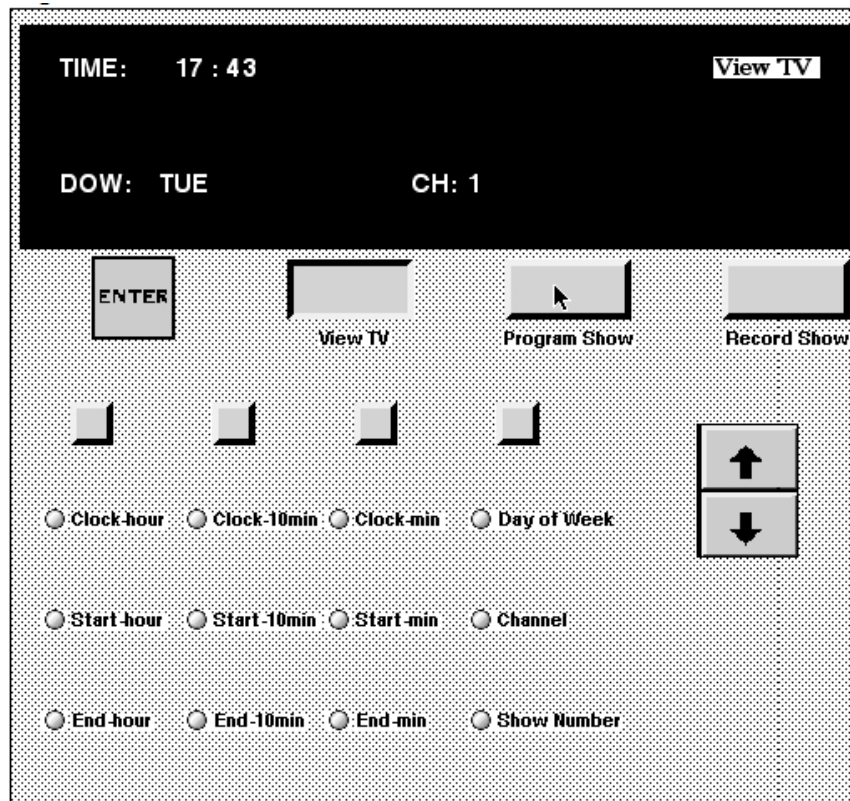
Figure 5 The display for the VCR task

Next, participants had to click the leftmost square button above the hour column

to "activate" the hour buttons. Following "activation" of the hour column, participants

had to select the "start-hour" radio button. Upon pressing the radio button, a number (1-

24) was displayed in the black box at the top of the interface. The participant then had to

click the up or down arrow keys until they reached the correct start-hour on the index

card. After selecting the correct hour, participants had to press the "Enter" button. Upon

pressing this button, the start hour in the black box disappeared – indicating that the value

had been entered. Next, participants had to press the "End-Hour" radio button. Upon

pressing the button, a number (1-24) was displayed in the black box at the top of the

interface. Participants then used the arrow buttons to select their desired end hour and pressed the "Enter" button (causing the number to disappear).

After finishing with the hour column, participants had to "deactivate" the hour column by pressing the leftmost square button located above the hour column. Next participants had to click the square button above the 10-minute column in order to "activate" the corresponding buttons. The 10-minute column corresponds to the tens-digit of the minute to be entered, and the minute column corresponds to the ones-digit of the minute to be entered. This helped reduce needless scrolling through 60 numbers. Participants continued through both the 10-minute and minute column in the same fashion until the entire start and end time was entered. This process was repeated for inputting the "Day of Week" and "Channel" information. After entering all show information, the "Program Show" button had to be "deactivated" and the "Record Show" needed to be "activated" to finish the trial. If at any time during the trial, the participant tried to click a button, or take an action in a column that had not been properly "activated," the computer would beep, signifying that the participant was trying to make an illegal move.

The interruption task for experiment 1 was a pursuit tracking task also used by Cades et al. (2011). The tracking task required participants to track a randomly moving aircraft using the mouse cursor (Figure 6).

24
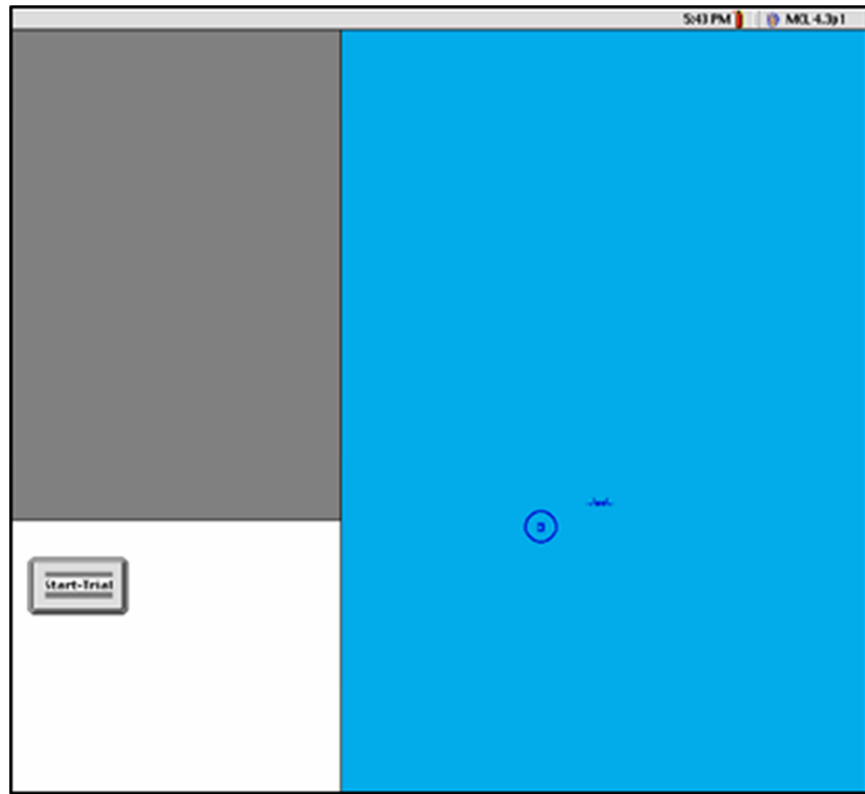
Figure 6 The display for the tracking task

The object of the task was to keep the mouse cursor (a circle with a dot) as close to the center of the aircraft as possible. Upon the onset of an interruption, the VCR task disappeared and the tracking task appeared. Upon resumption of the VCR task, the cursor was repositioned to its last position before the start of the interruption and the participant was allowed to resume the VCR task where they left off.

Design

This experiment was exploratory in nature and all participants served in the same condition to determine the nature of the decrease in resumption lag over time. Participants programed the VCR to record six television shows and experienced between

22 and 32 interruptions. Interruptions were triggered after a random number (7-17) of

clicks to reduce expectation and some participants were more efficient than others; thus,

the number of interruptions a participant received was slightly variable.

Measures and Analyses

The primary measure for this experiment was resumption lag, as described above.

However, whether participants resumed the primary task at the same place prior to being

interrupted was noted. Accuracy of the VCR task was determined by the number of

correct/incorrect values entered; accuracy of the tracking task was calculated using root

mean square error (RMSE). Inter-action interval, the average time between clicks on the

VCR task, was also calculated.

For all model fitting, the Bayesian Information Criterion (BIC) was used to

determine which model fit the data the best. When using BIC, the lowest number denotes

the best fit.Like r-squared, BIC compares how closely a proposed model fits a set of data.

However, unlike r-squared, BIC controls for the complexity of the model. R-squared will

almost always indicate that a more complex model fits the data better than a simpler

model, which leads to overfitting. The Bayesian Information Criterion penalizes a model

for being more complex. This ensures that any added model complexity is justified by a

sizeable increase to model fit.

Constant, linear, quadratic, power, and Boltzmann sigmoidal models based on the

hypothesized data distributions in Figure 4 were fit to the resumption lag data. The

constant model served as an important baseline for the Bayesian Information Criterion. If

the constant model fit the data better than the other proposed models, this would suggest

none of the models fit the data very well. If the proposed models fit better than the constant model, this indicates that the models are better at explaining the variance that the mean. The linear model corresponds to the outcome in Figure 4.B and the Boltzmann sigmoidal model corresponds to the outcome in Figure 4.C. Recall, that it was unclear whether the mechanisms behind outcome A from Figure 4 suggested a power function or quadratic function, so both were included. This resulted in five different models being fit to the data.

Apparatus

The experiment used a Macintosh G4 computer with a 17 inch CRT monitor. Television show information to be entered into the VCR was printed on 3 x 5 inch index cards posted to the left of the computer monitor. All index cards were posted to the side of the computer screen at the beginning of a trial. This allowed participants to program all shows sequentially without having to take a break while waiting for the experimenter to post a new card.

Participants

Sixty-two (16 men, 46 women) George Mason University undergraduate students participated in this experiment in exchange for course credit. Average participant age was 20.7 years; age ranged from 18 to 39 (SD = 3.2) years old. All participants had normal, or corrected to normal vision.

Procedure

All participants were run one at a time. Experimenters first demonstrated how to program the VCR to record a television show while the participants watched. After participants understood the task, they practiced programming two television shows without interruption. After the participants finished practicing the VCR task, the experimenter introduced them to the tracking task. Participants then begin a 60 second practice trial of the tracking task. Next, participants were instructed on how the VCR task would periodically interrupt them with the tracking task. Unlike other studies, participants did not have the opportunity to practice being interrupted before the start of the experimental trails. This ensured that participants had no prior experience being interrupted by this task or any ability to learn how to better cope with this interruption. This concluded the practice portion of the experiment.

In the experimental session, participants programed the VCR to record 6 television shows. While they were recording these shows, they were periodically interrupted with the tracking task. The tracking task lasted 10 seconds, and then participants were redirected back to the VCR programming task. After participants finished programming the VCR to record 6 television shows, the experiment was completed.

Results

Data from four participants were not collected due to technical difficulties. Data from two participants were thrown out because instructions were not followed. The remaining participants experienced 1468 interruptions, with a mean of 26.2 interruptions

per participant. Of the 1468 total interruptions, 48 (3.3%) were removed because they occurred after the end of the task. This happened when a participant finished entering the information for the final show, but had not yet pressed "Record Show" when the interruption occurred. Because of the way the program was set up, the resumption lag following this interruption was unable to be collected. Pearson correlations indicated that participants tended to make more errors during the beginning of the experiment compared to the end of the experiment (r = -.33, p= .052). Further, a Welsh two sample t-test confirmed that participants were slower to resume when they resumed the task incorrectly (M= 3.60 s) compared to correctly (M= 3.12 s, t= -2.25, p=.027). Therefore, all incorrect resumptions (95, 6.5% of all resumptions) were removed in an effort to ensure that any reduction in the resumption lag was not due to participants' increased ability to correctly resume after repeated interruptions.

Each participant received a different number of interruptions, because interruptions were triggered based on a random number of clicks, and some participants were less efficient at completing the task than others – leading to a greater number of interruptions. Several participants were able to complete all six trials with as few as 22 interruptions, while one participant experienced 32 interruptions across all six trials. Because participants who experienced more interruptions, also tended to be less efficient, any resumption lags for interruption numbers 23 or higher were upwardly biased by these participants. Therefore, only resumption lags for interruptions 1 through 22 were used in the following analyses.

Finally, resumption lag data were collapsed across participants using the median for each interruption number rather than the mean. Because resumption lags take on a positively skewed distribution due to a floor effect (similar to reaction time data), median values allowed analyses to not be overly affected by the elongated tails of the distributions.

VCR Task Performance

Performance on the VCR task ranged from 77% of information entered correctly to 100% of information entered correctly with a mean accuracy of 94% (SD= 4.1%). The correlation between Trial number and Accuracy was .01 (p=.53) indicating that participants did not get more accurate at performing the VCR task over repeated exposures.

Interruption Task Performance

The tracking task served as the interruption task in this experiment. Performance was calculated using the root mean square error (RMSE) with values logged at 3 hertz. Participants' RMSE across all interruptions ranged from17.5 pixels to 46.6 pixels with a mean RMSE of 26.0 pixels (SD= 5.04); it was not necessary to remove any data due to poor performance (average RMSE greater than 100 pixels). The correlation between Trial number and RMSE was -.17 indicating that participants did not get more accurate at performing the tracking task over repeated exposures. This suggests that the training was successful.

Inter-Action Interval Data

The inter-action interval is calculated as the average time between clicks for the VCR task, excluding the resumption lag. It is important to measure the inter-action interval because it serves as a baseline for how much of the reduction in resumption lag is due to general increased speed of completing the task. If the inter-action interval decreases at the same rate as the resumption lag, this would suggest that the only thing driving the reduction in resumption lag is general improvement in performing the task. If however, the slope of the resumption lag is greater than that of the inter-action interval, this suggests that there is a unique process occurring that is specific to resuming from an interruption.

In the following analysis, an adjusted inter-action interval was used to account for any general learning of the VCR task, independent of the interruption lag. Typically, the time between clicks would be calculated and then averaged together for each trial, however in the VCR task there were two main kinds of clicks. First, there were clicks that constitute being a sub-task or sub-sub-task in and of themselves. This included most clicks during the task. However, there were also clicks that were only part of a sub-task. This occurred when a participant was clicking the up or down arrows repeatedly, to get to a desired number. Because of this repetitive clicking, the time between these clicks tended to be very short, making the inter-click-interval artificially short. Therefore, the adjusted inter-click interval did not include repetitive clicking.

31
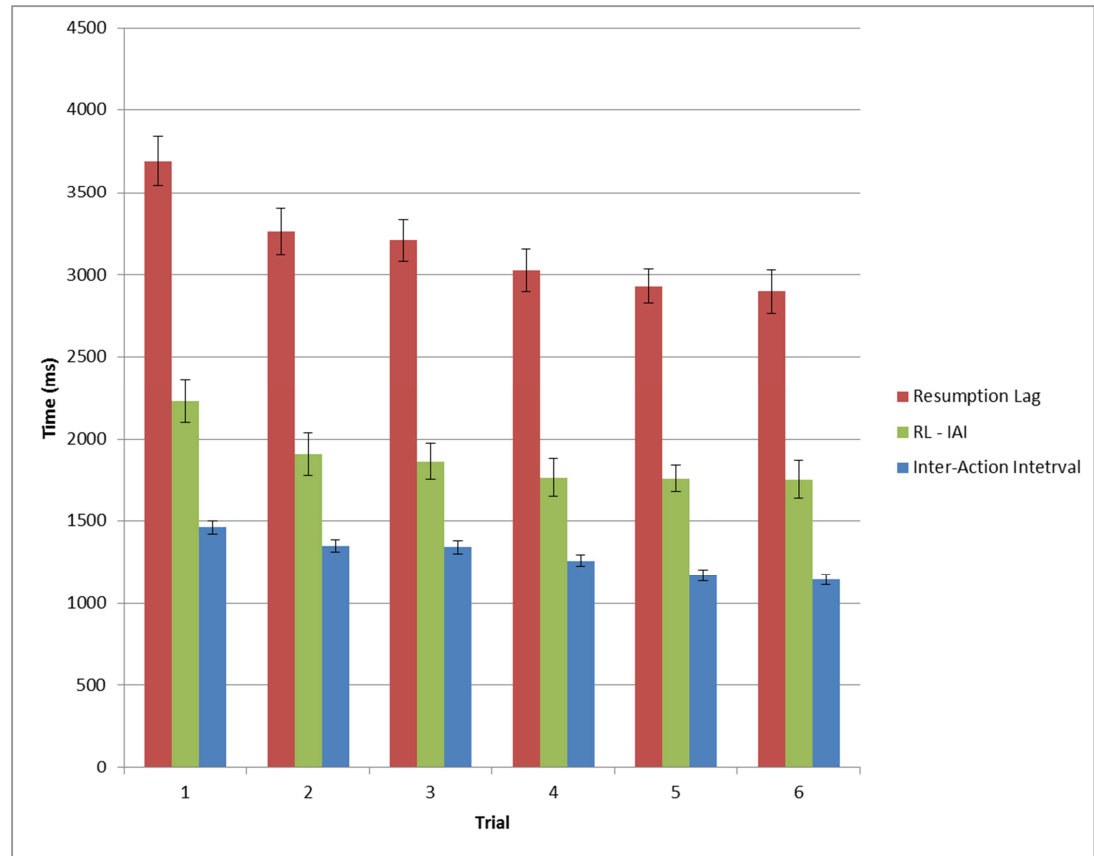
Figure 7. Average Resumption Lag, Inter-Action Interval, and Resumption Lag Subtracted by Inter-Action Interval (RL-ICI) by Trial for Experiment 1.

In an effort to filter out any reduction in the resumption lag that could be due to overall VCR task improvement, two analyses were performed . First, a repeated measures analysis of variance (ANOVA) using a Greehhouse-Geisser correction for sphericity indicated that there were significant within ($F= 7.70$, $df= 4.1$, $p < .001$) and between subjects ($F= 1796$, $df=1$, $p < .001$) effects for resumption lag. This indicated that the resumption lag was getting reliably smaller over repeated exposures. Next, each participant's average inter-click interval was subtracted from their average resumption

lag for each trial of the experiment. These differences were then run in a repeated measures ANOVA, which indicated that there were still significant within (F= 3.35, df= 4.3, p = .009) and between subjects (F= 704, df=1, p < .001) effects (see Figure 7). These analyses indicate that the reduction in resumption lag is reliable, even when general learning of the task is accounted for.

Resumption Lag Performance

Participants resumed from the interruption task as quickly as .58 seconds and as long as 16.5 seconds, with a mean resumption lag of 3.14 seconds (SD =1.79 s). Because resumption lags were positively skewed, median values were used instead of means for each level of interruption exposure.

Equations of best fit based on the hypotheses in Figure 4 were then fit to the data and evaluated using the Bayesian Information Criterion (BIC) fit statistic (see Figure 8). The first equation fit to the data was a constant model (y = mean(x)). Although this model was not predicted by any of the theories described above, it was included to serve as a baseline. If none of the models tested had a BIC value less than the constant model, this would suggest that there is no discernible pattern in the data. The BIC for the constant model was 319.9 and r^2 was 0.0 (the constant model serves as the baseline for r-squared).
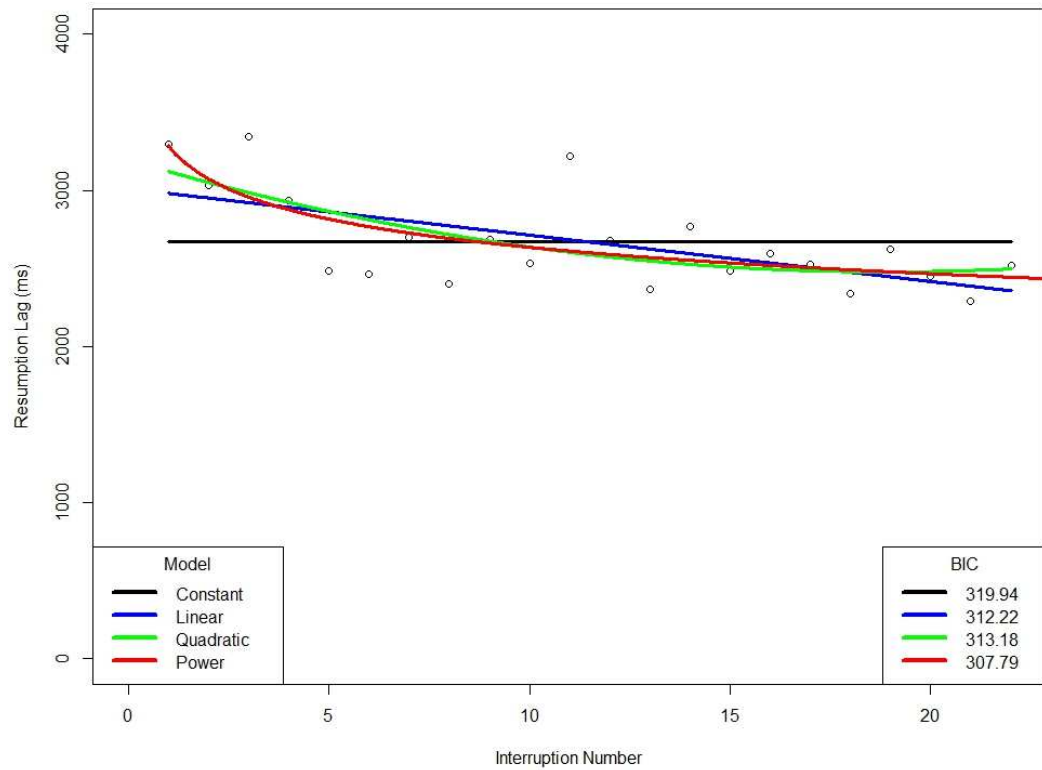
Figure 8 Hypothesized models fit to study 1 data

The linear model (y = ax + b) was tested next, because it was the second most parsimonious. It corresponds to outcome B in Figure 4, and was hypothesized to fit best by the Memory for Goals encoding mechanism as well as Threaded Cognition's problem state mechanism. The BIC for the linear model was 312.2 and r^2 was .39.

Outcome A in Figure 4 could be achieved by either a quadratic or power function, depending on how quickly learning occurs. Therefore, both models were fitted to the data to test Outcome A. The quadratic model (y = ax^2 + bx + c) had a BIC of 313.18 and an r^2 of .44 while the power function (y = ax^b) yielded a BIC of 307.79 (r^2 could not be

calculated because the power function is nonlinear). Finally, a Boltzmann sigmoidal model, which corresponds to outcome C of Figure 4 was fit to the data. This model represents the data distribution that would be expected based on the long term working memory theory. Unfortunately, the fit of this model was so poor that an equation of best fit would not converge. Therefore, estimates of fit could not be calculated.

Resumption Lag Modeling using Random Effects

In the above analyses, aggregated data were analyzed because they provided the best model fit, however it is important to understand whether aggregation is appropriate. It is possible that the models fit to the aggregated data don't actually fit any one participant's data very well. In an effort to test this possibility, a series of models were fit to the data: a random intercept model, fixed intercept with fixed linear slope, random intercept with fixed linear slope, and finally random intercept and independent linear slope.

The random intercept model fixes slope at zero and estimates each participant at their overall mean. If this model fit best, it would suggest that there are significant individual differences in resumption lag, but no reduction in resumption lag over repeated exposures. The fixed intercept with fixed linear slope model creates a single slope and intercept estimate for the entire dataset. This is the type of model that was used in the main analyses for the resumption lag data. If this model fit the data the best, it would suggest that the main analyses fairly represented all participants' data – suggesting that there were limited individual differences for average resumption lag as well as the rate of reduction in resumption lag. The next model used a random intercept with a fixed linear

slope. This model estimates an intercept for each participant, but only estimates a single slope that applies to all participants. If this model fit the data the best, it would suggest that even though all participants had a different starting point for resumption lag, they all experienced about the same reduction in resumption lag across repeated exposures. Finally, the random intercept with independent linear slope model was calculated. This model estimates an individual intercept and slope for each participant. If this model fit the data the best, it would suggest that not only are some participants initially faster at resuming to the primary task than others, participants also tend to have a unique rate of reduction in resumption lag over time.

The random intercept model (BIC = 19827) was first compared to the fixed intercept with fixed linear slope model (BIC = 19871) and the random intercept model fit the data better. This suggests that participants exhibited individual differences and the aggregated data may not have been as representative of any one individual participant as initially thought. The next comparison made was between the random intercept model (BIC = 19827) and the random intercept with fixed linear slope model (BIC = 19803). The random intercept with fixed linear slope model fit better, suggesting that adding a slope to the model instead of only estimating each participant at their individual mean explains more variance. This indicates that there was a reliable reduction in resumption lag over time. Finally, the random intercept with fixed linear slope model (BIC = 19803) fit better than the random intercept with independent linear slope model (BIC = 19810), suggesting that the rate of reduction in resumption lag over repeated exposures is similar across participants.

In the above analyses, the best fitting model was the random intercept with fixed linear slope model. This suggests that the average length of resumption lags varied between participants, but all participants' resumption lags tended to get faster at the same rate. Unfortunately, the model used in the main resumption lag analysis (fixed intercept with fixed linear slope) was not a good fit. This suggests that the models that were fit to the aggregated data may not fit the individual-level participant data very well.

To test this possibility, all hypothesized models (Constant, Linear, Quadratic, Power and Boltzmann Sigmoidal) were also fit to each participant's individual resumption lag data. These analyses used all resumption lag data from each individual participant and determined which of the above models best fit that participant's data using the BIC. Unfortunately, participants' individual resumption lag data were rather variable. Thus we decided to fit the models to the mean of each participant's resumption lags for each of the six trials. This resulted in six data points for each participant, with each of these data points being the average off all resumption lags for its respective trial. This helped smooth out much of the variability and better estimated what type of reduction in resumption lag participants were generally exhibiting over time. However, even after using the average resumption lag data for each trial, the power function and Boltzmann Sigmoidal model could not be fit to most participants' data. Therefore, these models were removed from this amalysis.

The results indicated that the constant model best fit 53.6% of all participants' trial mean resumption lag data, the linear model best fit 17.9% of all participants' trial mean resumption lag data, and the quadratic model best fit 28.6% of all participants' trial

mean resumption lag data. Although averaging resumption lags across trial helped smooth out some of the variance in the data, there were still a large proportion of participants whose data best fit the constant model. These results need to be considered when interpreting the aggregated data from the main resumption lag analyses – the aggregated data may not be sufficient for understanding the nature of the reduction in resumption lag that any one participant experienced.

Discussion

The goal of Experiment 1 was to determine the shape of the distribution of the resumption lags over time. By understanding the nature of the reduction in resumption lag, it may be possible to provide evidence for or against the three main theories of interruptions: Memory for Goals, Threaded Cognition, and Long Term Working Memory. Three potential patterns for the reduction in resumption lag were suggested by the three theories: a quadratic or power function, a linear function, and a Boltzmann sigmoidal function (see Figure 4). The encoding and base level activation mechanisms suggested by Memory for Goals predicted that the reduction in resumption lag over repeated exposures would be best fit by either a quadratic, power or linear function. The problem state mechanism suggested by Threaded Cognition predicted that the reduction in resumption lag over repeated exposures would be best fit by a linear or possibly Boltzmann sigmoidal function, and Long Term Working Memory theory suggested that the reduction in resumption lag would only be best fit by a Boltzmann sigmoidal function.

The pattern of the aggregate data in Experiment 1 was best fit by the power function (BIC = 308) followed by linear (BIC=312) and quadratic (BIC =313) models. This provides support for the encodingand base level activation mechanisms that are shared by Memory for Goals and Threaded Cognition. The data are not clear regarding the problem state mechanism proposed by Threaded Cognition. The best fitting model was the power function. The problem state mechanism of Threaded Cognition suggests that this type of function would not be likely. However, the next best fitting model was the linear function, which the problem state mechanism predicted.

These data distributions provide a novel way to evaluate theories of interrupted task performance. Although the distributions of aggregated data help us understand what is generally happening over time, it is also important to understand whether the observed distribution is homogeneous across participants. This is especially important for testing the Boltzmann sigmoidal model. If each participant's point of inflection fell in a somewhat different place along the distribution, the best fitting function would likely be linear. To test for this, each participant's data were individually fit to each of the proposed models.

The results from this analysis indicated that individual participants' data were most likely to be best fit by the constant model. This suggests that estimating slope for individual participant data tended to not explain much additional variance. This could have been caused by excess within-participant variance. If there was too much variance in the participants' individual data, the noise could outweigh the signal, and would cause

the BIC to better fit the less complex models. Therefore, changes were made in Experiment 2 in order to reduce within-participant variance.

The Boltzmann sigmoidal model fit the individual participant-level data so poorly that a model fit equation was not able to converge. On the surface, this appears to be strong evidence against Long Term Working Memory theory. However, because the individual participant data were variable and the Boltzmann sigmoidal model is sensitive to variability, the individual participant analyses may have unfairly discounted the Boltzmann sigmoidal model and Long Term Working Memory theory. Although this limitation is not likely to have affected the findings in this set of experiments, it must be considered when interpreting the conclusions made.

In summary, Experiment 1 analyzed the resumption lag data over repeated exposures in a way that no other experiment has to date. Past studies have compared mean resumption lag for each session, where each session consists of 10 to 12 interruptions. This experiment did not collapse across sessions; instead, it examined each resumption lag individually – evaluating the entire distribution of resumption lags. This is important because the theories of interrupted task performance each provide different mechanisms for how the reduction in resumption lag occurs. Moreover, each of these mechanisms would produce a distinct distribution. By comparing these distributions, we are better able to discern what mechanisms are driving the reduction in resumption lag over time. Experiment 1 suggested that the mechanisms proposed by Memory for Goals and Threaded Cognition could be accurate, while the mechanism proposed by Long Term

Working Memory Theory is not likely to be driving the reduction in resumption lag over

repeated exposures.

EXPERIMENT 2

Rationale

The distributions observed in Experiment 1 suggested that the Memory for Goals and Threaded Cognition models provide plausible mechanisms for the reduction in resumption lag over repeated exposures. However, Experiment 1 was exploratory in nature and was not able to specifically determine which model was better at explaining the reduction in resumption lag over repeated exposures. Thus, the goal of Experiment 2 was to focus on the problem state mechanism suggested by Salvucci and Taatgen's (2010) Threaded Cognition.

Salvucci and Taatgen's Threaded Cognition theory of learning states that participants experience interference when both a primary task and interruption task require a problem state. This interference leads to poor task performance and increased time to resume from an interruption. At first, beginners store their knowledge of task procedures in declarative memory, requiring a problem state. After becoming intermediate-to-expert at the task, the knowledge of the task procedures moves from declarative memory to procedural memory – no longer requiring a problem state for the task. However, if a procedural task requires retaining a piece of information in memory, it will always require a problem state – even if a person is an expert at the task.

In Experiment 1, the primary task was procedural and did not require the participant to hold anything in memory. Thus, according to the Threaded Cognition account of learning, the reduction in resumption lag in Experiment 1 would have been due to this transition of primary task procedural information from declarative to procedural memory.

Experiment 2 used a procedural task that could be manipulated so that it either always required a problem state (holding information in memory) or did not. Both Threaded Cognition and Memory for Goals models predicted a reduction in the transient problem state condition as seen in Experiment 1. Threaded Cognition explains the reduction in resumption lag in the transient condition as a conversion of the primary task requirements from declarative to procedural memory – thereby removing the problem state requirement for the primary task. Therefore, in the constant problem state condition, Threaded Cognition predicts no reduction in resumption lag because both the primary task and interruption task required a problem state for the duration of the experiment – causing interference (See Figure 9 for hypotheses).
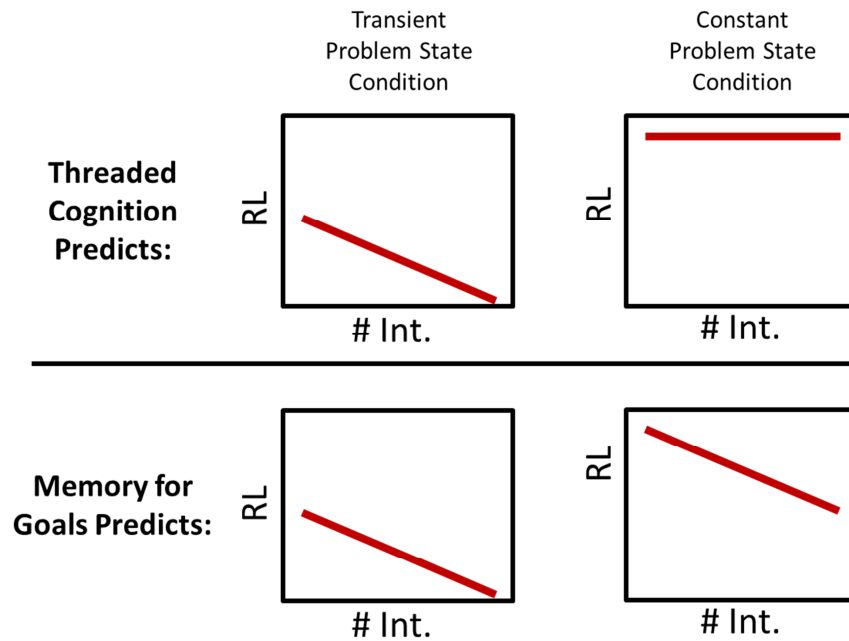
Figure 9 Patterns in resumption lag as hypothesized by Threaded Cognition and Memory for Goals.

Method

Tasks

The primary task was the same as Experiment 1 with a few changes. First, interruptions were triggered based on time (randomly chosen between 10 and 35 seconds) rather than number of clicks. This change served two purposes. First, participants in Experiment 1 reported getting very frustrated after being interrupted multiple times while trying to click the up or down arrow repeatedly to select a time or day. Secondly, this equalized the probability of an interruption occurring between all aspects of the primary task. In Experiment 1, a disproportionate number of interruptions occurred while clicking the up or down arrows.

The other change made to the primary task was the removal of all cues after resuming from an interruption. In Experiment 1, after an interruption, the cursor was moved back to where it was before the start of the interruption, and the column selection buttons and radio buttons remained selected (see Figure 5) – all giving participants cues as to where they left off. In Experiment 2, when a participant resumed from the interruption, the mouse was not reset and all column selection buttons and radio buttons were deselected to remove all cues.

In this experiment, it was important to have an interruption that required a problem state. Therefore, the interruption used was an n-back task. This task was chosen because the task always requires a problem state; that is, participants always had to store items in memory. Upon interruption from the primary task, participants were presented numbers (1-9) at a rate of one every 1.43 seconds. This presentation rate was determined by pilot testing and allowed for exactly 6 numbers to be presented (with 5 responses) during the 10 second interruption. These numbers were presented audibly with synthetic speech. Upon onset of a new number, participants were required to determine whether that number was smaller or larger than the number that was presented one stimulus previous. Participants made their selection using the mouse cursor and clicked lower or higher, respectively. Participants were automatically switched back to the primary task after 10 seconds of the interruption task.

Design

In this Experiment, there were two conditions. The first condition used the same primary task as Experiment 1, except for the noted changes. This was the condition where

transition from declarative to procedural memory is possible. The other condition was designed to induce a problem state that would not dissipate with practice. In this condition, the primary task stayed mostly similar to Experiment 1, except participants were required to memorize two pieces of information (Channel and Day of Week) about the show to be recorded rather than having all pieces of information freely available on a note card.

Participants programed the VCR to record six shows and experienced between 13 and 31 interruptions. Because interruptions were triggered every 10 to 35 seconds, the number of interruptions a participant experienced was variable.

Measures and Analysis

Measures for the primary task were the same as the previous experiment with one exception. Because the task was changed slightly, and all cues were removed upon resumption, participants had to make three clicks to resume what they were doing compared to a single click in Experiment 1. Therefore, resumption lag was defined as the time it took from the end of the interruption to the point where the participant had completely resumed the task (3 clicks). Because the interruption task was the n-back task instead of the tracking task, both response time and accuracy were measured. Similar to Experiment 1, observed distributions were compared to hypothesized models, and were analyzed using fit statistics.

Apparatus

All equipment was the same as Experiment 1.

Participants

Seventy-eight (20 men, 58 women) George Mason University undergraduate students participated in this experiment in exchange for course credit. Average participant age was 23.8 years, and ranged from 19 to 56 (SD = 6.38) years old. All participants had normal, or corrected to normal vision and hearing.

Procedure

The procedure was the same as Experiment 1; however, participants completed the n-back task instead of the tracking task, and in the problem state condition, participants were required to memorize some of the show information rather than having it freely available. Unlike Experiment 1, participants only had the information to program one show at a time. This served two purposes: first, it allowed participants to memorize the required information in the problem state condition. Second, it helped reduce the fatigue and frustration some participants reported in Experiment 1.

Results

Data from six participants were thrown out because instructions were not followed. Three additional participants were not able to satisfactorily complete the training tasks after repeated attempts, so they were dismissed prior to data collection. The remaining 69 participants experienced 1389 interruptions, with a mean of 20.1 interruptions per participant. Of the 1389 total interruptions, 119 (8.6%) were removed because they occurred either before the task began or after it ended. Because interruptions were triggered on time rather than actions, several participants were slow to begin the VCR task and were interrupted before even beginning the task. An additional 16 (1.2%)

interruptions were removed because two participants stopped following instructions half way through the experiment. As in Experiment 1, all incorrect resumptions (86, 6.2% of all resumptions) were removed and median values were used.

Participants experienced between 13 and 31 interruptions while setting the VCR to record all 6 shows. Each participant received a different number of interruptions, because interruptions were triggered at quasi-random time intervals. Further, some participants took longer to complete the task than others – leading to a greater number of interruptions. Because participants who experienced more interruptions, also tended to be less efficient, any resumption lags for interruption numbers 14 or higher were upwardly biased by these participants. However, only a few participants were able to complete the experiment with fewer than 16 interruptions, and there was no discernible increase in resumption lag for resumption lags 14 through 16 (see Figure 10). Therefore, resumption lags 1 through 16 were used and any resumption lag 17 or higher was thrown out.

In the memory condition, participants were required to remember both the day of week and the channel number for the show that they were recording to ensure that they had a persistent problem state throughout the experiment. However, participants were not always able to remember this information. If participants could not remember both pieces of information, the resulting resumption lags were removed from the analysis. Therefore, an additional 46 (3.3%) resumption lags were removed from the analysis.
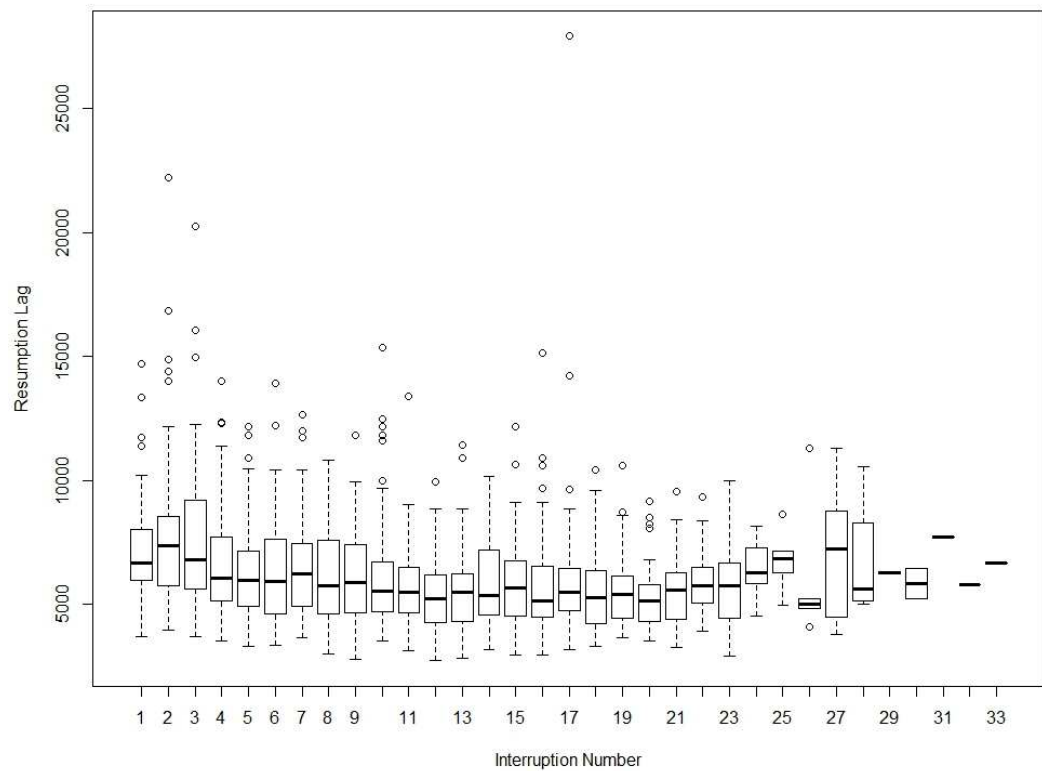
Figure 10. Distribution of resumption lags by interruption number for Experiment 2

VCR Task Performance

Performance on the VCR task ranged from 68% of information entered correctly to 100% of information entered correctly with a mean accuracy of 96% (SD= 5.7%). The correlation between Trial number and Accuracy was .06 (p=.002). Although this is a significant correlation, it is not meaningful, as it explains less than one half of one percent of the total variance.

Interruption Task Performance

The n-back task served as the interruption task in this experiment. Participants' accuracy ranged from 32% to 100% with an average accuracy of 90% (SD = 12.8%). All resumption lags where the corresponding n-back performance was less than 75% were removed to ensure that participants were actively engaged in the interruption task. This removed 155 resumption lags (11.2% of all resumption lags) from the data. Of the 155 resumption lags removed, 63 were from the control condition and 92 were removed from the memory condition.

Inter-Action Interval Data

The inter-action interval was calculated the same way as in experiment 1 and was winsorized at two standard deviations over the mean to reduce the effect of outliers. First, a mixed effects analysis of variance (ANOVA) indicated that there were significant within (F= 40.18, df= 5, p < .001) and between subjects (F= 3897, df=1, p < .001) effects for resumption lag. Resumption lags in both conditions dropped as trial number increased, and the large effect between participants indicates that participants significantly varied from one another.
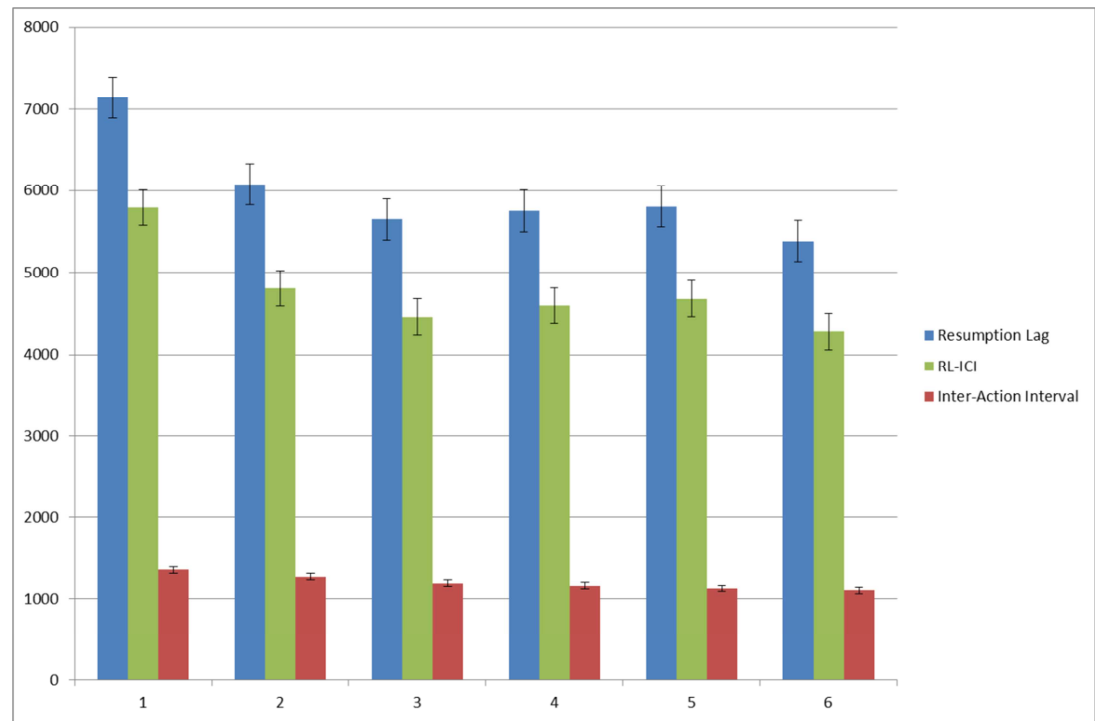
Figure 11. Resumption Lag, Inter-Action Interval and Resumption Lag subtracted by Inter-Action Interval (RL-ICI) by Trial Number for Experiment 2

To ensure that the reduction in resumption lag was due to a unique process involving the resumption lag and not simply improving at the primary task, the difference between the resumption lag and inter-action interval was calculated for each trial, for every participant. A mixed effects ANOVA indicated that there were still significant within (F= 31.4, df= 5, p < .001) and between subjects (F= 3573, df=1, p < .001) effects for this difference score. This suggests resumption lag is getting shorter over repeated exposures, even after accounting for general task learning.

Resumption Lag Performance

Participants' resumption lags ranged from 2.72 seconds to 27.9 seconds with a mean of 6.26 seconds (SD = 2.25 seconds). Because the data were positively skewed, median values were used instead of means for each level of interruption exposure.

Threaded Cognition suggests that if participants must maintain a constant problem state throughout the primary task, there will be no reduction in resumption lag across repeated exposures, while Memory for Goals suggests there should still be a reduction (see Figure 9). In order to test these hypotheses, a control condition without a constant problem state and a memory condition with a constant problem state were compared. Figure 12 shows the control condition with both constant (y = average(x)) and linear (y = ax + b) models fit to the data. As predicted by both theories, the linear model (BIC = 235) fits the data better than the constant model (BIC = 256). An analysis of variance comparing the residual sum of squares between the two models confirmed that the linear model fit better than the constant model (F=48,9, p<.001).
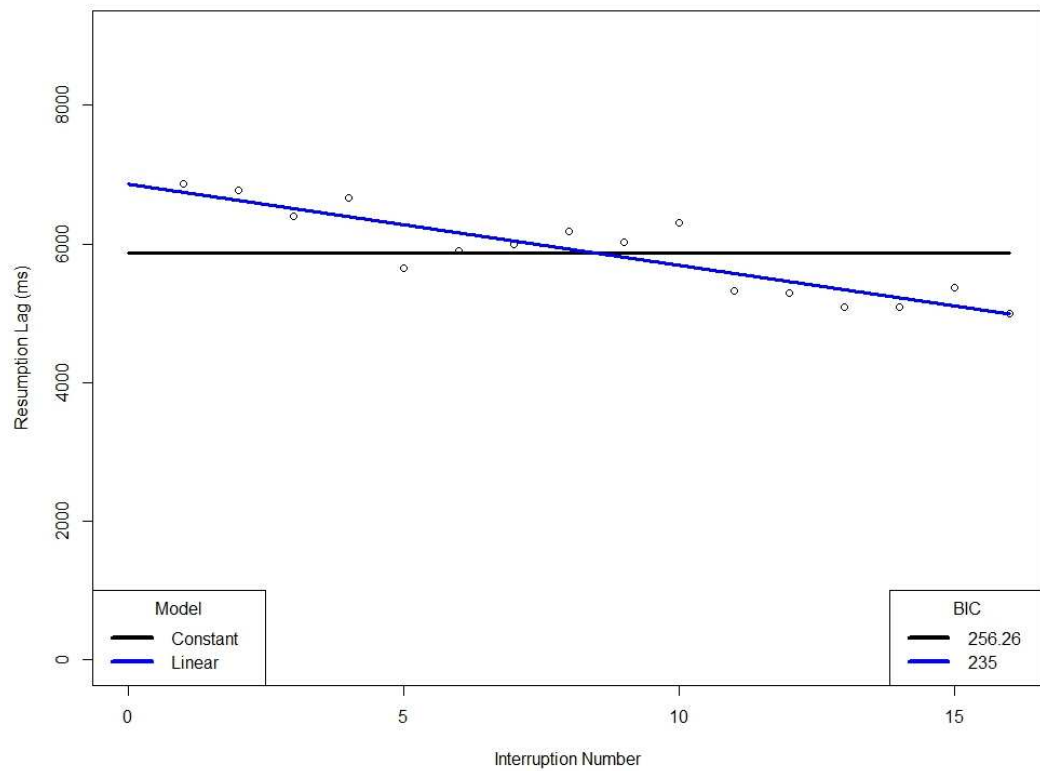
Figure 12 Median resumption lag by interruption number for the control condition

Next, both the constant and linear models were fit to the memory condition data.

Contrary to Threaded Cognition's problem state explanation, the linear model (BIC =

250) fit the data better than the constant model (BIC = 256; see Figure 13). An analysis of

variance comparing residual sum of squares confirmed that the linear model fit better

than the constant model (F= 10.7, p = .006).
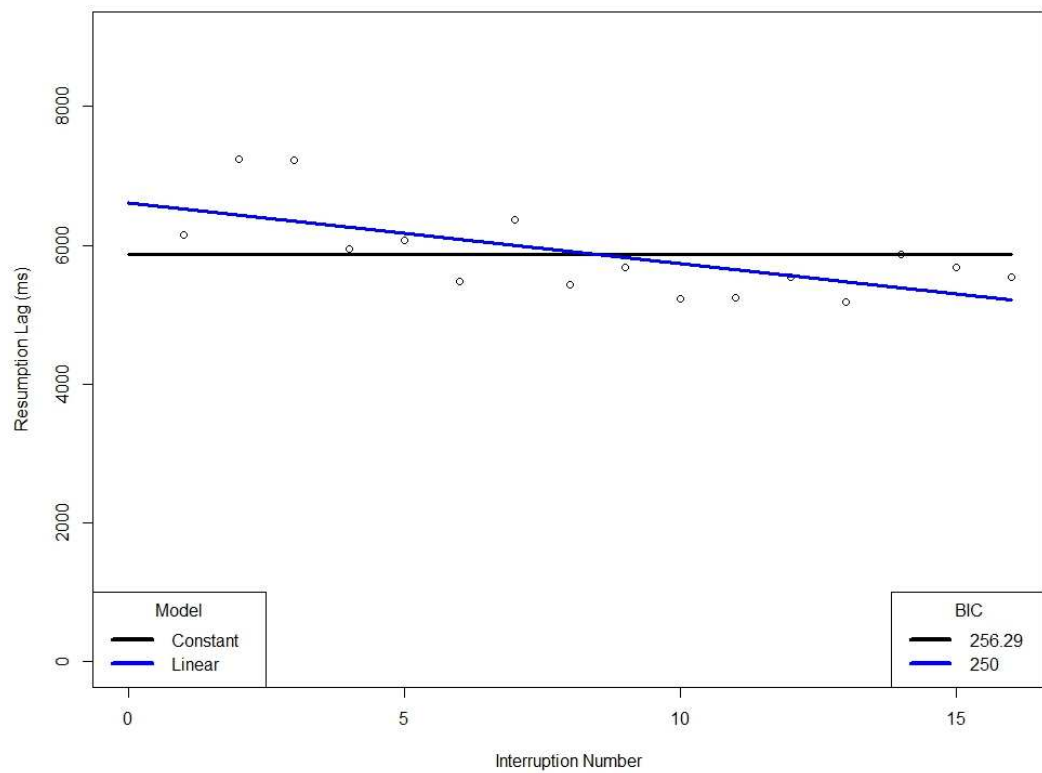
Figure 13 Median resumption lag by interruption number for the memory condition

Finally, the data were compared between the memory and control conditions using a two way analysis of variance, with condition and interruption number as independent variables and resumption lag as the dependent variable. There was a significant main effect for interruption number showing that resumption lag decreases as the number of interruptions experienced increases (F=41.9, p < .001). There was no main effect for condition, nor was there a significant interaction (see Figure 14).
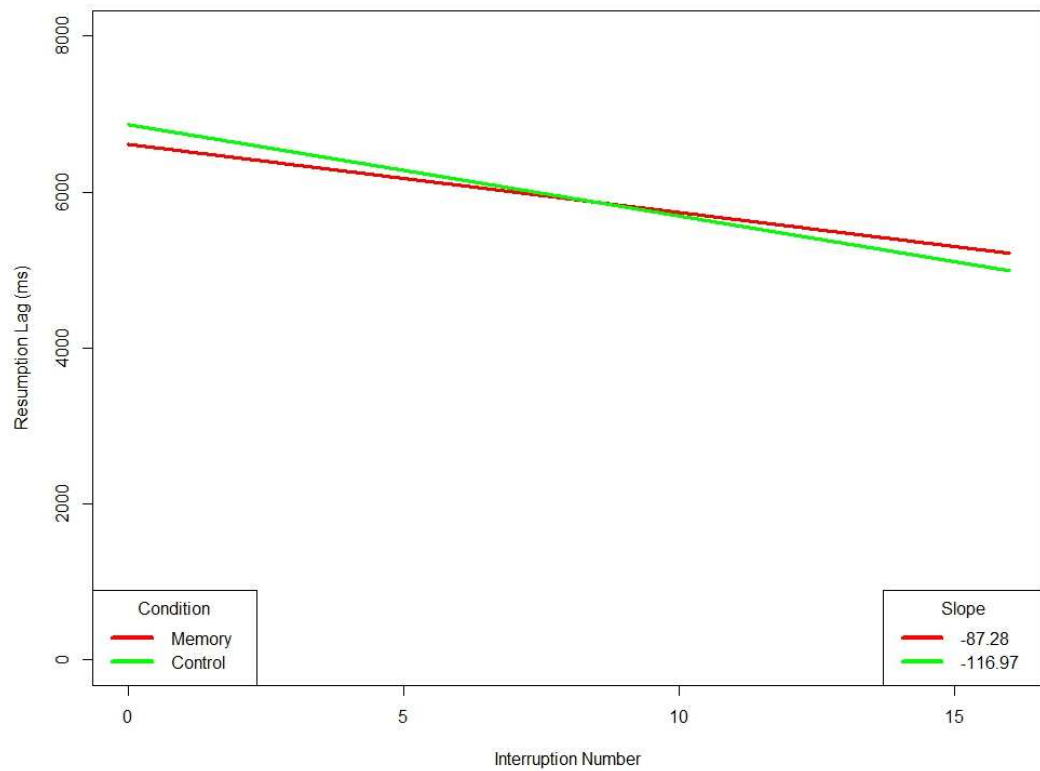
Figure 14 Comparison of memory and control conditions

Resumption Lag Model Fitting

Several changes were made between Experiment 1 and Experiment 2 to remove cues and make resumption lags less variable. One of the weaknesses of Experiment 1 was that the models that were fitted to the aggregated data did not fit individual participants' data very well. Because the control condition in Experiment 2 was identical to Experiment 1 with only the above noted changes, the hypothesized models from Experiment 1(Figure 4) were rerun using the Experiment 2 control data. Figure 15 shows

that the linear model fit the data the best (BIC = 235), followed by the quadratic

(BIC=237.7) and power models (BIC=239.7).
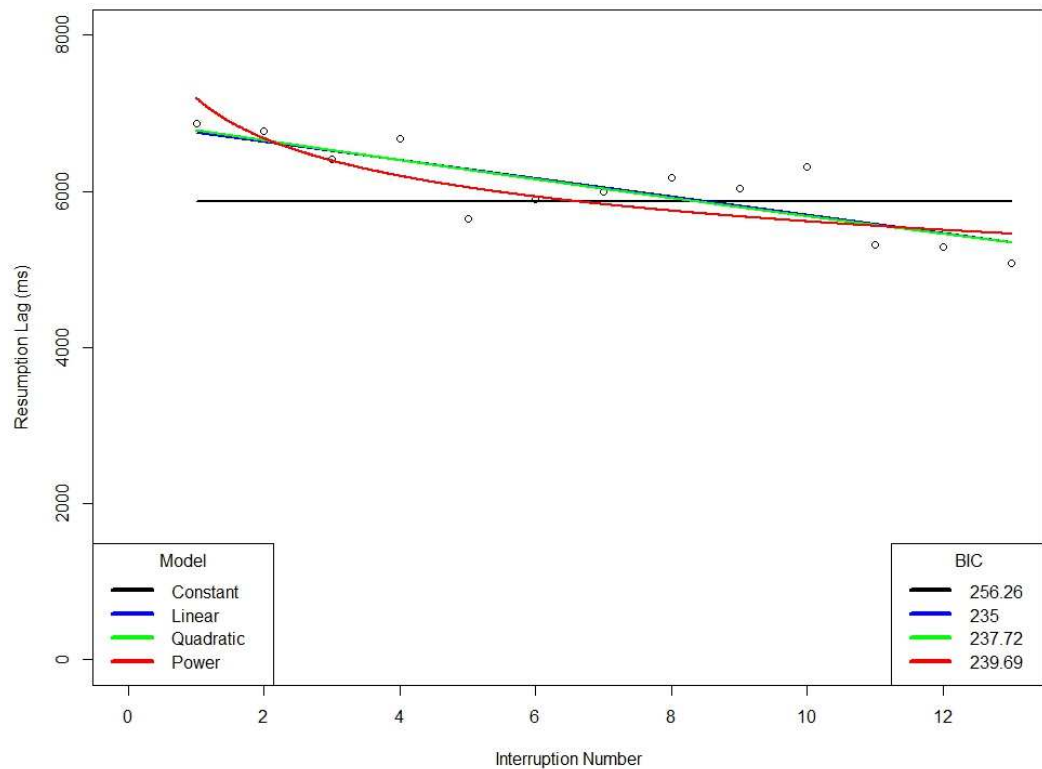
Figure 15 Model fitting for experiment 2 control condition

Compared to Experiment 1, the same general resumption lag distribution was

observed, with a few notable changes. First, in Experiment 1, resumption lags were

measured differently and were shorter on average than those in Experiment 2. Second,

the power function fit the data better than the linear model in Experiment 1. In

Experiment 2, the linear model fit better. Finally, all models fit the data better in Experiment 2 (best fit, BIC = 235) than in Experiment 1 (best fit, BIC = 308). This difference is most likely driven by increased variability in Experiment 1 due to some participants taking advantage of the available cues, while other participants did not notice, or take advantage of these cues.

Resumption Lag Modeling using Random Effects

Aggregated data were used in Experiment 2 because they provided the best model fit, however it is important to understand whether aggregation was appropriate. Similar to Experiment 1, a series of models were fit to the data: a random intercept model, fixed intercept with fixed linear slope, random intercept with fixed linear slope, and finally random intercept and independent linear slope. As in Experiment 1, the random intercept model with fixed slope was the best fitting model. This suggests that the average time to resume varied between participants, but all participants' resumption lags tended to get faster at the same rate.

It is therefore important to better understand if the models fit to the aggregated data accurately represented individual participants. Because the individual-level participant data were highly variable, the proposed models were fit to the mean of each participant's resumption lags for each of the six trials. This resulted in six data points for each participant, with each of these data points being the average of all resumption lags for its respective trial. This allowed for fitting the models to a participant's overall trend, while smoothing out some of the within-participant variance.

The results indicated that the quadratic model was the best fit to participants' trial mean resumption lag data for 47.8% of all participants. The constant model best fit 27.5% of all participants. Finally, the linear model best fit 24.6% of participants' resumption lag data. This means that most (72.4%) participants experienced a reduction (either quadratic or linear) in resumption lag. Because the aggregated data were also best fit by the quadratic and linear models, this provided evidence that the aggregate data was a fairly good representation of the overall resumption lag data.

Discussion

There were two goals for Experiment 2. The primary goal was to evaluate Threaded Cognition's problem state mechanism for explaining the reduction in resumption lag over repeated exposures. The secondary goal was to fit the models proposed in Experiment 1 to the Experiment 2 control data. To examine the problem state mechanism, we manipulated the need for participants to carry a problem state during the primary task.

Experiment 2 found that resumption lag in the constant problem state condition did not remain constant over repeated exposures. Instead, it got shorter at about the same rate as the control condition. This finding did not support the problem state mechanism suggested by Salvucci and Taatgen (2010). However, it is important to understand any potential limitations of this experiment before any conclusions should be made. First, participants were required to remember the channel and time that the show was to be recorded. It is possible that requiring participants to remember only these two pieces of information was too easy of a task, and did not fully engage the problem state resource

for the primary task. If this was the case, it could have been possible to have a reduction in resumption lag over repeated exposures in the memory condition, but perhaps not at the same rate as the control condition. Although the difference in slope between the memory and control conditions was not significant, the slope in the control condition (-117 ms per resumption) was higher than the memory condition (-87 ms per resumption), lending some credence to this theory.

The second goal of Experiment 2 was to replicate the findings in Experiment 1 using the Experiment 2 control condition. There were features of Experiment 1 that were hypothesized to contribute to excess variance and changes were made in Experiment 2 to address this concern. In this analysis, the linear and quadratic models best fit the aggregated data. Importantly, when the proposed models were fit to individual participant data, 72.4% of participants' data were best fit by the quadratic or linear models, with only 27.5% of participants' data best fit by the constant model. This suggests that the experimental manipulations were successful in reducing within subject variability, and that the aggregated data represented individual participant data fairly well.

In summary, Experiment 2 was able to use the experimental approach developed in Experiment 1 and evaluate the problem state mechanism suggested by Salvucci and Taatgen (2010) in Threaded Cognition. The findings in this experiment did not provide evidence in favor of the problem state mechanism. The control condition from Experiment 2 was also used to rerun the analyses from Experiment 1. The results from this analysis supported the conclusions made in Experiment 1 and provided better evidence against Long Term Working Memory Theory.

EXPERIMENT 3

Rationale

Experiment 2 found that resumption lags tended to get smaller over repeated exposures in the constant problem state condition. This did not support Salvucci and Taatgen's (2010) problem state mechanism proposed as part of Threaded Cognition. Although Experiment 2 did not explicitly test the Memory for Goals theory of interrupted task performance, it did provide some evidence in its favor. Namely, Memory for Goals suggests that there should be a reduction in resumption lag, regardless the number of items being held in memory. Therefore, experiment 3 was designed to explicitly test the Memory for Goals theory of interrupted task performance.

Experiment 3 was based on work by Trafton et al. (2003). They found that participants in an immediate interruption condition showed a reduction in resumption lag across repeated exposures while participants who were warned of an impending interruption had a relatively constant, and much shorter resumption lag (see Figure 3). One possible explanation for these results is that participants who were warned of in impending interruption were initially better able to encode the primary task, which led to faster resumptions than participants without a warning. However, over time, participants in the no warning condition were better at speeded encoding - completing the encoding task while completing the interruption task. While these participants initially took longer

to resume from interruptions, they were able to get faster as they learned to both encode and complete the primary task at the same time – leading to a reduction in resumption lag over repeated exposures.

In Experiment 3, participants either experienced no alarm, or experienced an interruption alarm for the first half of the trials and then experienced no interruption alarm for the second half of trials. If the reduction in resumption lag was due to participants' improved ability to encode the primary task while performing the interruption task, no improvement in encoding should occur during the period where an alarm is used. This is hypothesized because an interruption alarm allows for encoding to occur before an interruption begins, and participants are not required to encode while performing a secondary task. Because ability to encode did not improve during the alarm condition, the speeded encoding mechanism of Memory for Goals would suggest that participants' resumption lags should immediately but temporarily increase when they switch to the no alarm condition, (see Figure 16.C). In the condition where participants receive no alarm throughout the whole experiment, a distribution similar to Figure 16.A is expected.
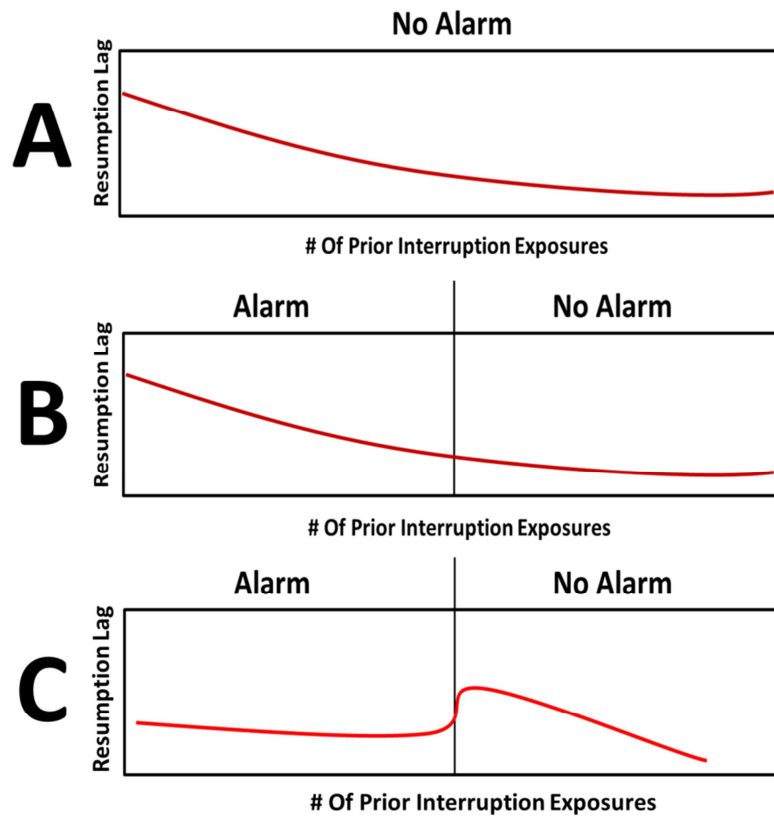
Figure 16. Hypothesized pattern of reduction in resumption lag in the No Alarm condition as hypothesized by both Threaded Cognition and Memory for Goals (A). Hypothesized Patterns in resumption lag in the Alarm condition as hypothesized by Threaded Cognition (B) and Memory for Goals (C).

Method

Tasks

In this experiment, there were two conditions. The No Alarm condition used the same version of the primary task and interruption task from Experiment 2 with no other changes. The Alarm condition also used the same task, except the first three trials had a brief auditory alarm that sounded five seconds before the start of an interruption. In the last three trials, there was no interruption alarm. Initially, participants were generically

instructed to "prepare for the interruption" when they heard the interruption alarm. Most participants chose to continue working until they either found "good" place to stop, or where automatically transferred to the interruption task. After initial analysis of the data, it was apparent that the interruption alarm was not aiding in faster resumption as it did in Trafton et al (2003) or Hodgetts & Jones (2006). Therefore, new participants were run in this condition and instructed to stop working on the primary task immediately after they heard the interruption alarm and spend the five seconds remembering where in the primary task they left off. This had the added benefit of ensuring that participants were indeed encoding during the interruption lag and not exclusively during the interruption. Only data from the second set of alarm condition participants were used in this experiment.

Design

This experiment was a between-subjects design with two conditions: Alarm and No Alarm. In the alarm condition, participants programed the VCR to record three shows with the aid of an interruption alarm followed by three shows without the aid of an interruption alarm. In the no alarm condition, participants were interrupted without warning for all six trials. In both conditions, participants programmed the VCR to record six shows and experienced between 13 and 39 total interruptions. Except for the alarm, all other aspects of the primary task and interruption task were the same as Experiment 2.

Measures and Analysis

  Measures for the primary task and n-back task were the same as in Experiment 2. As in Experiment 2, resumption lag was defined as the time it took from the end of the interruption to the point where the participant had completely resumed the task (3 clicks).

Apparatus

  All equipment was the same as Experiments 1 and 2.

Participants

Seventy-six (23 men, 53 women) George Mason University students participated in this experiment in exchange for either course credit or $15 compensation (not including participants removed from the first cohort of the alarm condition). Average participant age was 21.1 years, and ranged from 18 to 53 (SD = 5.3) years old. All participants had normal, or corrected to normal vision and hearing.

Procedure

  The training procedure was the same as Experiment 2, except that participants were not required to memorize the show information on their note card. Participants programed one show at a time until they programmed all six shows. In the alarm condition, participants were warned of an impending interruption while programming the first three shows. They were then told that they would no longer have an interruption alarm, and then they programmed the last three shows without an interruption. In the No Alarm condition, participants programmed all six shows without the aid of an alarm.

Results

Participants experienced a total of 1482 interruptions with a mean of 19.5 interruptions per participant. Of the 1482 interruptions, 90 (6.1%) were removed because the interruption occurred either before the start, or after the end of the primary task. This occurred because the alarm was triggered on a random time interval and some participants did not immediately begin the primary task. As in Experiments 1 and 2, all incorrect resumptions (122; 8.2%) were removed from the data. Because interruptions were triggered after participants had been working on the primary task for between 10 and 35 seconds, and because some participants were faster than others, participants experienced a variable number of interruptions. Participants who were able to complete the task very quickly only experienced 13 interruptions, while one participant experienced 39 interruptions (The mean number of interruptions experienced was 19.5). Because participants who experienced more interruptions also tended to take longer to complete the task, any resumption lags for interruption numbers 14 or higher were upwardly biased by these participants.

However, only a few participants were able to complete the experiment with fewer than 16 interruptions, and there was no discernible increase in resumption lag for resumption lags 14 through 16 (see Figure 17). Therefore, resumption lags for interruptions 1 through 16 were used in the following analyses. As in Experiments 1 and 2, medians were used instead of means because of the positively skewed nature of response time data.
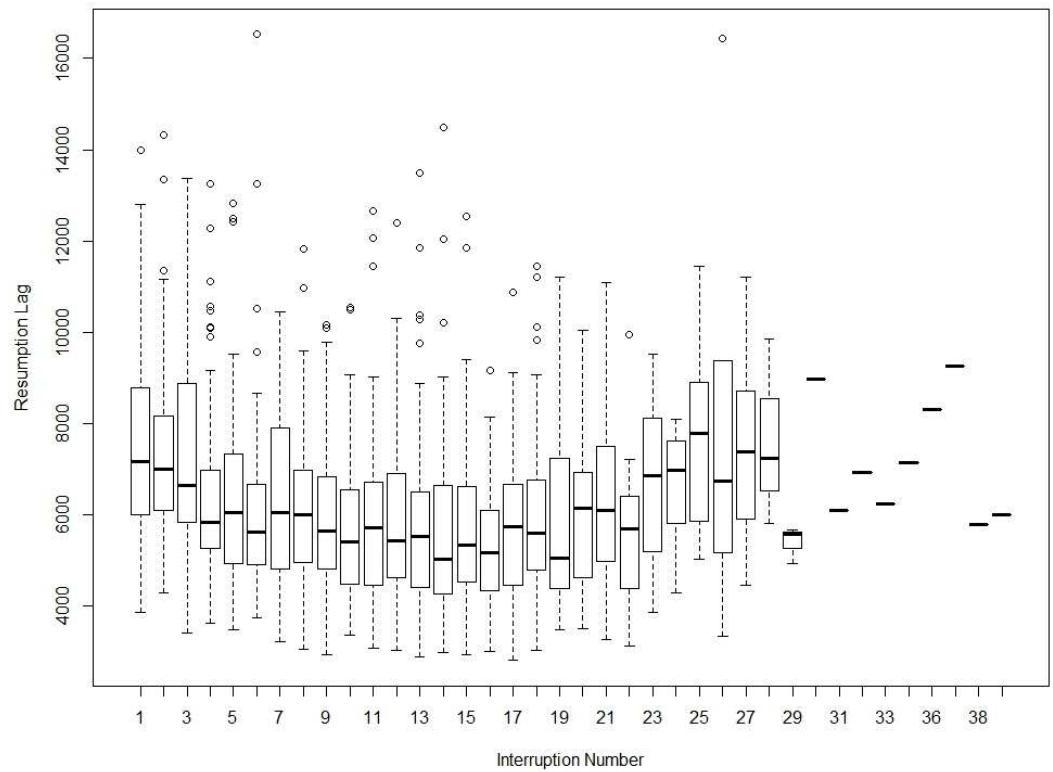
Figure 17. Distribution of Experiment 3 Data by Interruption Number

VCR Task Performance

Average performance on the VCR task ranged from 79% of information entered correctly to 100% of information entered correctly with a mean accuracy of 96% (SD= 4.7%).

Interruption Task Performance

As in Experiment 2, the n-back task served as the interruption task. Participants' task performance ranged from 54% correct to 100% correct, with an average performance of 91% correct. The average n-back performance in the alarm condition was 91% correct compared to 89% correct for the control condition. Any resumption lag for which the previous n-back performance was less than 75% correct was removed from the resumption lag analysis. This resulted in the removal 152 resumption lags (10.3% of all resumption lags) from the analysis – 61 from the alarm condition and 91 from the control condition.

Inter-Action Interval Data

The inter-action interval was calculated the same way as in Experiments 1 and 2 and, the data were winsorized at two standard deviations over the mean to reduce the effect of outliers. First, a mixed effects analysis of variance (ANOVA) indicated that there were significant within ($F = 41.42$, $df = 5$, $p < .001$) and between subjects ($F = 4126$, $df = 1$, $p < .001$) effects for resumption lag (see Figure 18). Resumption lags in both conditions dropped as trial number increased, and the large effect between participants indicates that participants significantly varied from one another.
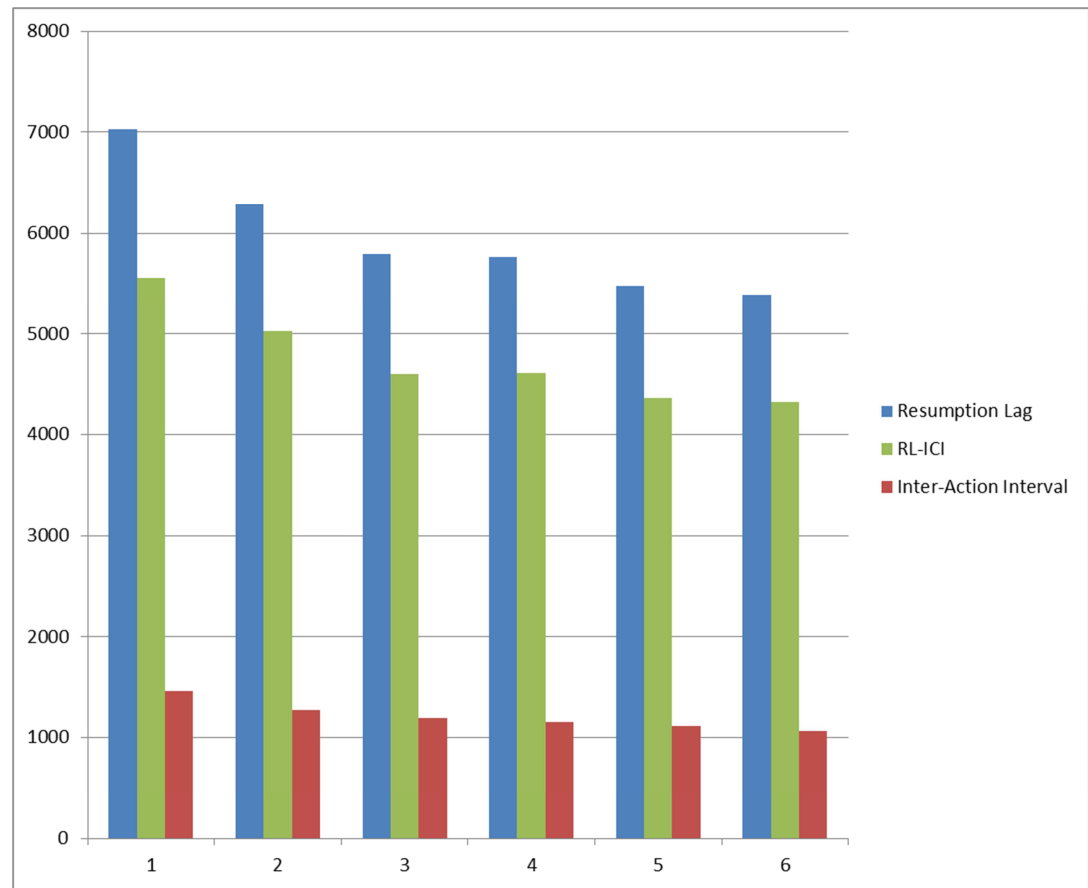
Figure 18. Resumption Lag, Inter-Action Interval and Resumption Lag Subtracted by Inter-Action Interval (RL-ICI) by Trial Number for Experiment 3

To ensure that the reduction in resumption lag was due to a unique process, and not simply improving at the primary task, difference of the resumption lag and inter-action interval was calculated for each trial, for each participant. A mixed effects ANOVA indicated that there were still significant within ($F = 28.4$, $df = 5$, $p < .001$) and between subjects ($F = 4011$, $df = 1$, $p < .001$) effects for this difference score (see Figure 18). This suggests that resumption lag is getting shorter over repeated exposures, even after accounting for general task learning.

Resumption Lag Performance

Participants' resumption lags ranged from 2.8 seconds to 16.5 seconds with an average of 6.2 seconds. As in the first two experiments, medians were used instead of means so that these positively skewed data did not affect our measure of central tendency.

In this experiment, an interruption alarm was used to help determine what was driving the reduction in resumption lag over time in the first two experiments. If the Memory for Goals speeded encoding mechanism is correct, and the reduction in resumption lag is due to improved or speeded encoding during the interruption, we would expect either a small decrease in or consistently lower resumption lag during the alarm condition, followed by an increase in resumption lag once the alarm is no longer present, followed by a consistent reduction in resumption lag (see Figure 16.C). There are two analyses that must be conducted to test this hypothesis: We must first make sure that the interruption alarm worked as it was intended. This will be accomplished by comparing the resumption lag for the first interruption in the alarm and no alarm condition. For the alarm to be successful, the alarm condition must have a shorter resumption lag than the no alarm condition. Second, in the alarm condition, we must compare the last resumption lag using an interruption alarm to the first resumption lag without an alarm. This will determine if the resumption lag gets longer as predicted by Memory for Goals.

First, we must compare the resumption lags for the alarm condition to the resumption lags for the no alarm condition for the first few exposures. This analysis is important, because it shows whether the interruption alarm was effective in reducing the

resumption lag by allowing participants to rehearse and encode before the start of the interruption. After removing data that were more than two standard deviations above the mean, a one tailed student's t-test indicated that the first resumption lag was significantly shorter in the alarm condition (m = 6.78 s) compared to the control condition (m = 7.60 ; t = -1.81, p = .038). This confirmed that the alarm condition was successful in reducing the resumption lag relative to the control condition (see Figure 19).
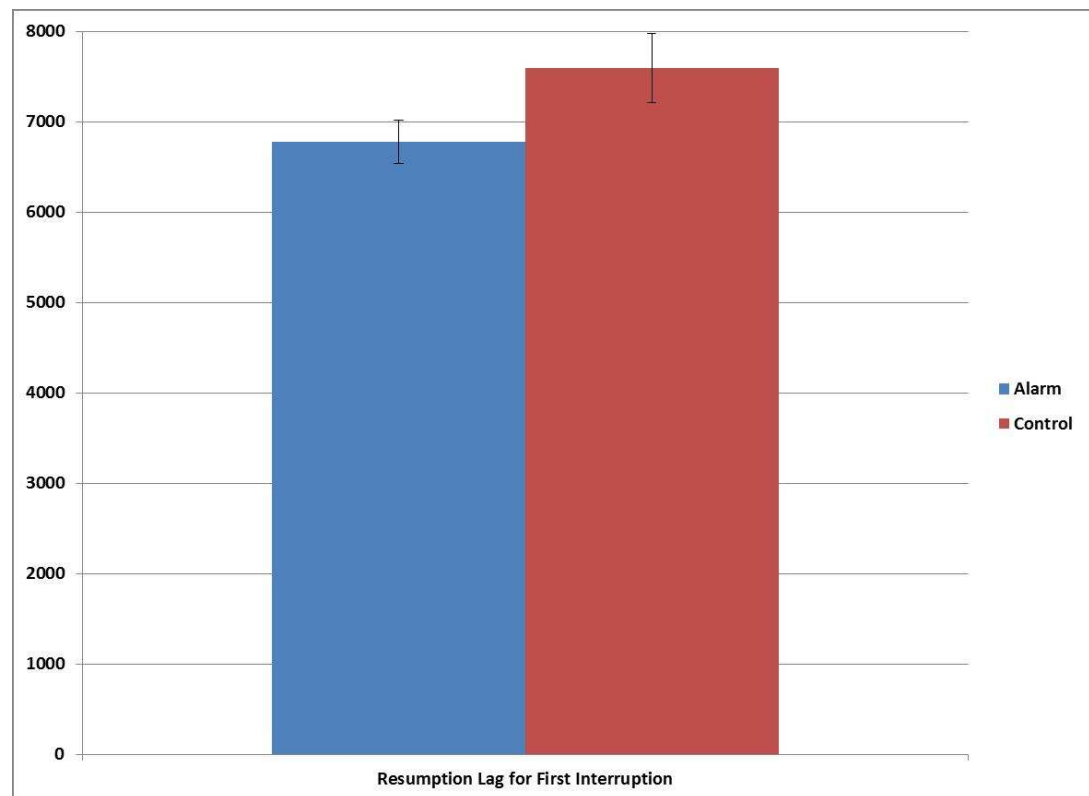
Figure 19. Comparison of first resumption lag across Alarm and Control conditions

Next, participants' resumption lags for the last interruption while using an alarm was compared to their first interruption without the alarm. Memory for Goals suggests that these values should be different, while Threaded Cognition suggests that there would be no difference (See Figure 16). A one tailed student's t-test indicated that there is no difference between participants last interruption with the alarm (interruption -1 in Figure 20; m = 5.67 s) and participant's first exposure to an interruption without an alarm (interruption 0 in Figure 20; m = 5.41 s, t = .89, p = .81). This provides support against the Memory for Goals speeded encoding explanation.
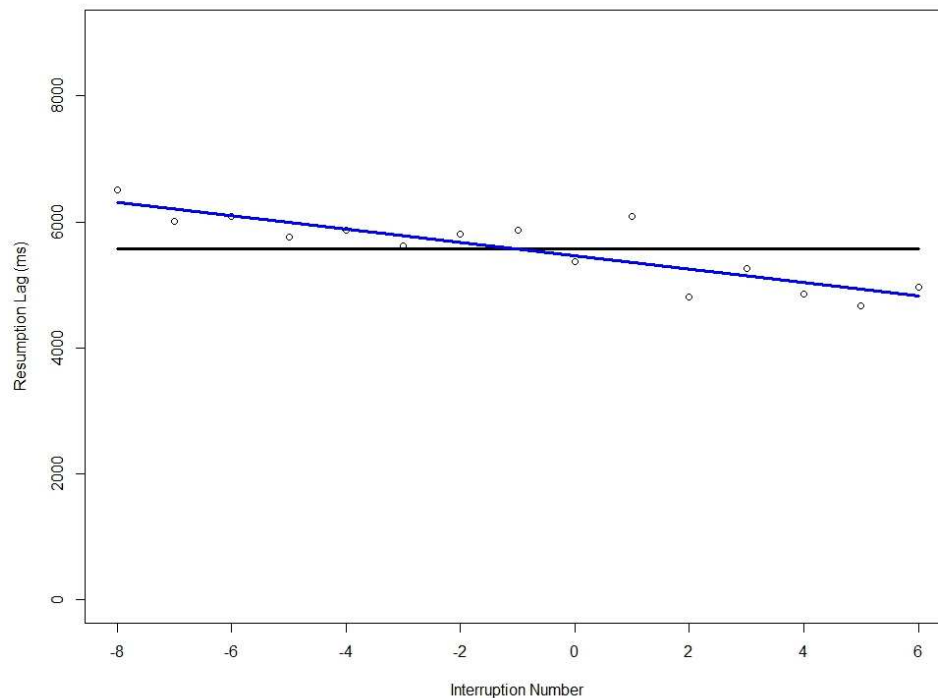
Figure 20. Reduction in resumption lag for Alarm condition

71

Discussion

The goal of Experiment 3 was to evaluate the speeded encoding mechanism derived by the Memory for Goals theory (Altmann and Trafton, 2002). To test this, participants in the alarm condition experienced an interruption alarm for the first 3 trials and then no interruption alarm for the last three trials. If increased ability for speeded encoding was driving the reduction in resumption lag over repeated exposures, we would expect a lower, but constant resumption lag for the first three trials of the experiment, followed by an immediate increase, and then steady reduction in resumption lag over repeated exposures in the second three trials (see Figure 16). In the first three trials, participants in the alarm condition should have had sufficient time to encode the primary task before being interrupted – leading to a consistently fast resumption lag. In the second three trials, when the alarm was removed, an immediate increase in resumption lag was expected because participants' lack of experience dual tasking the encoding of the primary task with performing the interruption task.

The results indicated that there was no difference in resumption lag between the last interruption lag with an alarm and the first interruption lag without an alarm. These results were inconsistent with the hypothesized pattern of resumption lags. However, it is important to consider the limitations of this experiment. First, the reduction in resumption lag for the first three trials of the alarm condition was not significantly lower than the first three trials of the control condition. This suggests that the interruption alarm may not have had the same effect seen in Trafton et al. (2003). In their experiment, they found that resumption lag for the interruption alarm condition was almost 50% of the

resumption lag for the no interruption alarm condition, but this effect diminished over repeated trials (See Figure 3). In our experiment, the resumption lag for the very first interruption in the alarm condition was significantly lower than the first interruption in the no alarm condition; however this effect was not consistent for the duration of the first trial. Moreover, the resumption lag in the alarm condition in the Trafton et al. experiment remained relatively constant over repeated exposures. In our experiment, the resumption lag in the alarm condition got shorter over repeated exposures. Together, these differences suggest that the interruption alarm in Experiment 3 may not have been as effective as Trafton et al.'s (2003) interruption alarm.

It is possible that even though participants were required to stop working on the primary task and use the interruption lag for encoding, they may not have actually been encoding. A subset of participants in the alarm condition (23) were asked if they preferred having an alarm or no alarm; 65% indicated that they preferred the no alarm condition. Of those participants, all indicated that they preferred no alarm because it was "faster" or they "didn't have to wait around" for the interruption task to begin. If they were performing the task as instructed, they should have filled that time with rehearsal of the primary task and should not be waiting for the interruption to begin. In summary, even though support was not found for speeded encoding, there may have been other factors that contributed to these findings.

GENERAL DISCUSSION

In this set of experiments, three hypothesized mechanisms were compared in their ability to explain the reduction in resumption lag over repeated exposures: Creation of retrieval structures, problem state, and speeded encoding. Long Term Working Memory theory suggests that specialized retrieval structures are created in long term working memory and it is the eventual creation of these structures that causes the reduction in resumption lag over repeated exposures. Our studies have shown that this is not the mechanism driving the reduction in resumption lag. Threaded Cognition suggests that the reduction in resumption lag is driven by the loss of interference due to the removal of the main task problem state; however we did not find strong support for this mechanism. Finally, the speeded encoding mechanism suggested by Memory for Goals suggests that the reduction in resumption lag is caused by improving one's ability to encode while completing the interruption task; however, this mechanism was not supported by our studies (See Figure 21 for a summary of findings).

Memory for Goals / Threaded Cognition	Exp 1	Exp 2	Exp 3	Summary
Encoding	✓		O	O
Strengthening	✓			
Base level Act.	✓			
Threaded Cognition Problem State	✓ (Partial)	O		O
Long term working memory	✗			✗

✓ = Supported

✗ = Not Supported

O = Further Research Required

Figure 21. Summary of Findings

Creation of Retrieval Structures

The first experiment evaluated the distribution of resumption lags over repeated exposures to determine whether the shape of the distribution was consistent with that predicted by each mechanism. The mechanism suggested by Long Term Working Memory was the creation of retrieval structures in long term working memory; this mechanism would suggest that the distribution of resumption lags would follow a stepwise function. However, when a stepwise function (Boltzman sigmoidal model) was fit to the data, it fit so poorly that a function of best fit would not converge.

An examination of the aggregated data suggested that there was no obvious inflection point at which resumption lags became much faster. However, this pattern

could have arisen from individual participants following a perfect stepwise function, but with different inflection points. Therefore, the stepwise function was also fit to the individual participant-level distributions of resumption lags. The stepwise function was not able to converge on a function of best fit for most of the participant-level data and it was not the best fitting model for any participant when a solution did converge.

One interpretation of these two analyses could be that they provide strong evidence against the creation of retrieval structures mechanism. However, it is important to understand the limitations of fitting the stepwise model to the individual-level data. The stepwise function is very complex and the individual-level data have a high level of random variance due to participants paying closer attention during some trials than others. Therefore, even if the creation of retrieval structures was driving the reduction in resumption lag, it is unclear whether this analysis would be sensitive enough for the stepwise function to best fit the majority of participants' data.

With the relatively weak test of the creation of retrieval structures mechanism suggested by Long Term Working Memory, was it premature to rule this mechanism out? We believe it was justified for theoretical reasons. First, the theory suggests that a person remains slow at resuming from an interruption until they become an expert at the given task and their retrieval structure in long term working memory is fully created. At this point, the expert can now almost instantly encode task related information to this structure, and when they resume from an interruption, they will have near instant access to this information, leading to much faster resumption lags. Given that the experiment lasted roughly 30 minutes, it is highly improbable that participants became an expert

during this short time. Therefore, there could not have been sufficient time to allow for the creation of these retrieval structures. Nonetheless, our participants exhibited a reduction in resumption lags over the course of the experiment. Thus, we believe it is unlikely that the reductions we saw were caused by the creation of retrieval structures. This does not rule out, however, the possibility that this mechanism could cause a reduction in resumption lag in a task practiced over a period of many years.

Problem State

The next mechanism evaluated in this series of experiments was the problem state mechanism suggested by Threaded Cognition. Experiment 2 directly tested this mechanism by using a memory condition and a control condition. Because memorizing this information required declarative memory, even if the participant became an expert at the primary task, the primary task still required a problem state. This means that there should be interference throughout the entire experiment, and according to Salvucci and Taatgen (2010), participants' resumption lags in the memory condition should remain high, and not get shorter over repeated exposures. The results from Experiment 2 did not support this mechanism. Resumption lags in the memory condition got shorter over repeated exposures, even though both the primary task and interruption task required a problem state.

Although the resumption lags in the memory condition did not remain unchanged as predicted by the problem state mechanism, the slope of resumption lags over repeated exposures was less for the memory condition than it was in the control condition by 30 milliseconds per interruption. This difference between the slopes is not statistically

significant, however, it may be operationally significant – after 15 interruptions, participants in the control condition were resuming almost a half second faster than those in the memory condition. These results suggest it is possible that the memory task used was not difficult enough. Had the memory task been more difficult, it is possible that the slope for the memory condition would have been more in line with the problem state mechanism's predictions. Because the data were trending in the direction predicted by the problem state mechanism, we can conclude that our experiment did not find evidence for this mechanism; however, further research using a more difficult memory task is warranted in order to draw a more decisive conclusion.

Speeded Encoding

The goal of Experiment 3 was to explicitly test the encoding mechanism proposed by Memory for Goals. Unfortunately, it is difficult to empirically test the encoding mechanism as a whole because it is impossible to keep participants from encoding. Therefore, speeded encoding, the ability to encode while concurrently completing the interruption task, was evaluated. Trafton et al. (2003) found a reduction in resumption lag only after immediate interruptions, not after interruptions with an interruption lag. If improved speeded encoding is driving the reduction in resumption lag, then this finding could be due to participants' increased ability to encode while completing the interruption task.

In Experiment 3, participants in the alarm condition experienced an interruption alarm in the first half of the experiment, and then no interruption alarm for the second half of the experiment. In the first half of the experiment, the primary task did not need to

be encoded while completing the interruption task because the interruption alarm allowed for encoding prior to the onset of the interruption. Therefore, we expected that the speed with which the primary task would be encoded would remain relatively low, and not improve over the course of the first half of the experiment. In the second half of the experiment, no warning was provided. This means that participants were required to encode the primary task while performing the interruption task. The speeded encoding mechanism suggests that when participants are suddenly faced with no warning, performance should deteriorate since they would not have learned to encode the primary task quickly during the first half of the experiment.

The results from Experiment 3 did not support speeded encoding as participants continued to get faster at resuming throughout the entire experiment. Losing the interruption alarm did not increase their resumption lags, as predicted. This provides evidence that the speeded encoding mechanism is not driving the reduction in resumption lag that is seen over repeated exposures.

Conclusion

In this set of experiments the creation of retrieval structures, removal of problem state interference, and improvement in speeded encoding mechanisms were all tested for their ability to explain the reduction in resumption lag over repeated exposures. The creation of specialized retrieval structures mechanism suggested by Long Term Working Memory theory was not supported by our findings. The problem state mechanism suggested by Threaded Cognition was not supported by our data, however, it is possible that with a more difficult memory condition, support may have been found. Finally, the

speeded encoding mechanism derived from the Memory for Goals theory of interrupted task performance also was not supported by our findings.

Moving forward, future research should further evaluate the problem state mechanism proposed by Threaded Cognition by using a more difficult memory condition. Additionally, the general encoding mechanism derived from memory for goals should be further evaluated. Finally, while not tested in this set of experiments, the base-level activation mechanism derived from Memory for Goals should also be evaluated.

REFERENCES

REFERENCES

Altmann, E. M., & Trafton, J. G. (2002). Memory for goals: An activation-based model. *Cognitive science*, *26*(1), 39–83.

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An Integrated Theory of the Mind. *Psychological Review*, *111*(4), 1036–1060.

Anderson, J. R., & Lebiere, C. (1998). *The Atomic Components of Thought*. Psychology Press.

Anderson, J. R., Qin, Y., Sohn, M.-H., Stenger, V., & Carter, C. (2003). An information-processing model of the BOLD response in symbol manipulation tasks. *Psychonomic Bulletin & Review*, *10*(2), 241–261.

Anderson, John R. (2007). *How Can the Human Mind Occur in the Physical Universe?* Oxford University Press.

Bainbridge, L. (1984). Diagnostic skills in process operation. *Proceedings of the 1984 International Conference on Occupational Ergonomics*, *2*, 1–10.

Basex. (2006, January 9). Workplace interruptions cost US economy $588 bn a year. *Basex*. New York.

Borst, J. P., Taatgen, N. A., Stocco, A., & van Rijn, H. (2010). The Neural Correlates of Problem States: Testing fMRI Predictions of a Computational Model of Multitasking. (B. J. Harrison, Ed.)*PLoS ONE*, *5*(9), e12966.

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

Czerwinski, M., Horvitz, E., & Wilhite, S. (2004). A diary study of task switching and interruptions. *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 175–182).

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, *102*(2), 211–245.

Flynn, E. A., Barker, K. N., Gibson, J. T., Pearson, R. E., Berger, B. A., & Smith, L. A. (1999). Impact of Interruptions and Distractions on Dispensing Errors in an Ambulatory Care Pharmacy. *American Journal of Health-System Pharmacy*, *56*(13), 1319–1325.

Gray, W. D. (2000). The nature and processing of errors in interactive behavior. *Cognitive Science*, *24*(2), 205–248.

Hodgetts, H. M., & Jones, D. M. (2003). Interruptions in the Tower of London task: Can preparation minimise disruption? *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 47, pp. 1000–1004).

Hodgetts, H. M., & Jones, D. M. (2006). Contextual cues aid recovery from interruption: the role of associative activation. *Journal of Experimental Psychology: Learning, Memory, and Cognition; Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(5), 1120.

Latorella, K. A. (1998). Effects of Modality on Interrupted Flight Deck Performance: Implications for Data Link. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *42*(1), 87–91.

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

Salvucci, D. D., & Taatgen, N. A. (2008). Threaded cognition: An integrated theory of concurrent multitasking. *Psychological Review*, *115*(1), 101.

Salvucci, D. D., & Taatgen, N. A. (2010). *The Multitasking Mind*. Oxford University Press.

Santell, J. P. (2005). Medication errors: experience of the United States Pharmacopeia (USP). *Joint Commission Journal on Quality and Patient Safety / Joint Commission Resources*, *31*(2), 114–119, 61.

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

CURRICULUM VITAE

Erik T. Nelson was born in Overland Park, Kansas. He received his Bachelor of Science in Cognitive Psychology from Kansas University in 2008, and his Master of Arts in Psychology from George Mason University in 2010.